Engineering Materials

Amretashis Sengupta
Chandan Kumar Sarkar *Editors*

Introduction to Nano

Basics to Nanoscience and Nanotechnology

Springer

Engineering Materials

More information about this series at http://www.springer.com/series/4288

Amretashis Sengupta · Chandan Kumar Sarkar
Editors

Introduction to Nano

Basics to Nanoscience and Nanotechnology

Springer

Editors
Amretashis Sengupta
School of VLSI Technology
Indian Institute of Engineering Science and Technology
Shibpur
Howrah
India

Chandan Kumar Sarkar
Department of Electronics and Telecommunication Engineering
Jadavpur University
Kolkata
India

ISSN 1612-1317 ISSN 1868-1212 (electronic)
Engineering Materials
ISBN 978-3-662-47313-9 ISBN 978-3-662-47314-6 (eBook)
DOI 10.1007/978-3-662-47314-6

Library of Congress Control Number: 2015941119

Springer Heidelberg New York Dordrecht London
© Springer-Verlag Berlin Heidelberg 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media
(www.springer.com)

Dedicated to Sri Sri Kaibalyanath, my parents and my teachers

Amretashis Sengupta

Foreword

Nanotechnology looks very much on the course to kick-start a new era in science and technology as well as industry in decades to come.

The true potential of the nanoscale, if properly exploited, would significantly improve many aspects of human living for the better.

In this scenario, nanotechnology education will play a pivotal role in preparing industry and society to fully embrace this new paradigm.

The efforts of Sengupta and Sarkar in providing students with a useful resource on this topic under one cover for venturing into the world of nanotechnology are most commendable.

A good aspect of this book is that it does not require any prerequisite of detailed knowledge of physics and materials sciences, but rather endeavors to introduce these concepts in a self-contained manner.

In this sense, the book strikes a nice balance between the building blocks of nanoscale physics and advanced topics such as nanotransistors, memories, and interconnects.

I feel that this book will be an important read for the young nanotechnologists and early-stage researchers, especially those working in the field of nanoelectronics.

Prof. Stevenson Fung
Department of Physics
The University of Hong Kong
Pokfulam Road
Hong Kong
China

Preface

It is said that some of the biggest surprises often come in small packages such a statement fits perfectly in the case of nanotechnology. This interesting new field of science and technology is getting more and more intertwined with the thread of our daily lives by means of nanoscale devices in consumer electronics, nanomaterials, and composites in textiles, aerospace, automobile, rehabilitation aids, and various other products. Not only nanotechnology is employed in the high-tech products, but also it is actively pursuing pressing global problems such as safe drinking water, better crops, and cleaner energy.

As nanotechnology builds momentum to become the game changer of the twenty-first century, the study of nanoscience and technology as an interdisciplinary subject in academic institutions also holds a key to its success. People and ideas from various fields of natural and applied sciences and engineering can come together under the aegis of the nanotech wave to transform the future for the better.

In this wave, this book is but a tiny ripple to provide students new to the field of nanoscience and technology the basic skill sets and building blocks to form a better understanding of the subject. With a special focus on nanoelectronic devices, this book is likely to be useful for early-stage researchers working in the domain of nanoelectronics as well. The materials and chapters for this edited volume have been collected from various experts and researchers working actively in the concerned fields. We hope the students would find it a suitable companion as they take their first steps into the marvelous world of nanoscience and technology.

Finally, we would like to express our sincere thanks to all the contributors and authors, and colleagues without whose active participation it would not have been possible to bring out this book. Our special thanks to our students for their enthusiasm which motivated us in the first place to initialize such an effort.

Howrah and Kolkata
May 2015

Amretashis Sengupta
Chandan Kumar Sarkar

Contents

Introduction 1
Amretashis Sengupta and Chandan Kumar Sarkar

Fundamentals of Quantum Theory 7
Amretashis Sengupta

Basic Solid-State Physics and Crystallography 27
Amretashis Sengupta

Quantum Nanoengineering 59
Arkaprava Bhattacharyya and Vijay K. Arora

Quantum Effect on Properties of Nanomaterials 73
Chandan Kumar Ghosh

Advanced Characterization Techniques 113
Chandan Kumar Ghosh

Nanoscale MOSFET: MOS Transistor as Basic Building Block 145
Soumya Pandit

Applications of Nanotechnology in Next-Generation Nonvolatile Memories 173
Amretashis Sengupta, Bikash Sharma and Chandan Kumar Sarkar

Nanocrystalline Thin Film Gas Sensors 205
Amretashis Sengupta

Editors and Contributors

About the Editors

Dr. Amretashis Sengupta received his B.Sc. (Hons.) in physics in 2005 and M.Sc. (physics) with specialization in Electronics in 2007 from the University of North Bengal, India. He recieved his M.Tech. (Nano Science and Technology) degree in 2010 from Jadavpur University, India. He received the Ph.D. (Engg.) degree from Jadavpur University, Kolkata, India, in December 2012. He is a recipient of university medals for ranking 1st class with 1st in M.Sc. and M.Tech. examinations. He was awarded the DST INSPIRE Fellowship for his doctoral work at Jadavpur University and the DST Postdoctoral Fellowship in NanoScience and Technology for his postdoctoral work at Indian Institute of Science, Bangalore. He received the DST INSPIRE Faculty award from the Department of Science and Technology, Government of India, in 2013. He is currently an assistant professor at the School of VLSI Technology, Indian Institute of Engineering Science and Technology (IIEST), Shibpur, India. His research interests are first principles-based simulations on 2D materials, non-equilibrium Green's function (NEGF)-based nanoelectronic device simulation for Graphene and other alternate 2D channel material FET, and nanocrystal-based nonvolatile memories. He is a member of IEEE and a reviewer of various reputed international journals such as IEEE T-ED, Journal of Computational Electronics, and Journal of Physics and Chemistry of Solids and Silicon.

Prof. Chandan Kumar Sarkar received the M.Sc. degree in physics from the Aligarh Muslim University, Aligarh, India, in 1975; the Ph.D. degree from Calcutta University, Kolkata, India, in 1979; and the D.Phil. degree from the University of Oxford, Oxford, UK, in 1983. He was a Research Fellow of the Royal Commission for the Exhibition of 1851 at the Clarendon Laboratory, University of Oxford, from 1983 to 1985. He was also a Visiting Fellow of the Linkoping University, Linkoping, Sweden. He joined Jadavpur University, Kolkata, in 1987, as a reader in electronics and telecommunication engineering and became a professor and the Head of the Department of Physics, Bengal Engineering College (currently Bengal Engineering and Science University), Howrah, India, in 1996. Since 1999, he has been a professor with the Department of Electronics and Telecommunication Engineering, Jadavpur University. He has served as a Visiting Professor in many universities and has published around 100 papers in referred journals. Dr. Sarkar is the Chairman of the IEEE Kolkata Section and Vice-chair of the IEEE Electron Devices Society (EDS) Region 10 (Asia Pacific Region). He serves as a Distinguished lecturer of the IEEE EDS. Prof. Sarkar has also authored/edited various text and references such as Optoelectronics and Fiber Optics Communication (New Age International, 2001), VLSI Design and EDA Tools (Scitech 2011), and Technology Computer-Aided Design: Simulation for VLSI MOSFET (CRC Press 2013).

Contributors

Vijay K. Arora Faculty of Electrical Engineering, Department of Electronic and Computer Engineering, Universiti Teknologi Malaysia UTM, Skudai, Johor, Malaysia; Department of Electrical Engineering and Physics, Wilkes University, Wilkes-Barre, PA, USA

Arkaprava Bhattacharyya Faculty of Electrical Engineering, Department of Electronic and Computer Engineering, Universiti Teknologi Malaysia UTM, Skudai, Johor, Malaysia; School of Electrical and Electronics Engineering, SASTRA University, Thanjavur, Tamil Nadu, India

Chanda Kumar Ghosh School of Materials Science and Nanotechnology, Jadavpur University, Kolkata, West Bengal, India

Soumya Pandit Institute of Radio Physics and Electronics, Calcutta University, Kolkata, India

Chandan Kumar Sarkar Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India

Amretashis Sengupta School of VLSI Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India

Bikash Sharma Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India

Introduction

Amretashis Sengupta and Chandan Kumar Sarkar

Over the last 15–20 years there has been a lot of buzz surrounding the word "nanotechnology" [1–7]. A lot is being printed in the press and being aired in the media regarding how this new technology is changing the world around us with myriad applications. It is already touching the everyday lives of people in various avatars such as *nanoparticles* containing sunscreen lotions, dirt-resistant glass/paint with *nanocoatings*, the lab on a chip (also known as the micro testing and analyzing system; μ-TAS in short), and the various *nanocomposite* materials. Also with ongoing research, the *SMART* uniforms for soldiers, shape-shifting aircraft wings, self-repairing concrete, and swarms of intelligent *nanobots* may not be limited to science fiction movies in future. In the electronic devices field, the discovery of graphene (a two-dimensional allotrope of carbon) in 2004 has created massive enthusiasm owing to its superior electronic properties than any conventional semiconductor known to man. From all these developments, nanotechnology seems to be bridging the gap between creative imagination and reality [2–11].

So all this makes us wonder what exactly is this nanotechnology and how does it differ from the per say "pre-nano" technologies of yesteryears? The term *nano* simply denotes a scale of 10^{-9}, in terms of dimension (size) of materials the nanometer thus means 10^{-9} m. So the objects falling under the nanotechnology regime are likely to have one or more dimensions (length/width/height) in the nanometer scale. So the science and technology related to such objects, which could be ranging from simple clusters of molecules known as ultrafine or nanoparticles, to complex systems like nanoelectromechanical system (NEMS) to devices like the single electron transistors [4–11].

A. Sengupta (✉)
School of VLSI Technology, Indian Institute of Engineering Science and Technology,
Shibpur, Howrah 711103, India
e-mail: a.sengupta@vlsi.iiests.ac.in

C.K. Sarkar
Department of Electronics and Telecommunication Engineering, Jadavpur University,
Kolkata 700 032, India
e-mail: phyhod@yahoo.co.in

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_1

But there exists a very fundamental question—just how many nanometers make an object truly fall under the scope of nanotechnology? 1, 10, 100, 500 nm? and just how small/large an object qualifies as a "proper" nanomaterial? Usually, it is considered that objects having one of the dimensions below 100 nm should fall into this regime of nanomaterial. However, this definition is rather tricky because the true essence of nanotechnology lies not only in the dimension but equally (maybe more importantly) on the alteration of material properties at small dimensions. Thus, even a 200 nm object may well be considered as a "proper" nanomaterial if it indeed shows some significant departure from its material properties at bulk or say micrometer scales.

So *Nanotechnology is the science and technology of objects in the nanoscale, whose properties differ significantly from that of their constituent material {of the said object} at the macroscopic or even microscopic scale.*

1 What Makes Nano Work?

There's plenty of room at the bottom—Richard P. Feynman, NL.

Nobel Laureate R.P. Feynman's way back in 1959 had envisioned the extreme miniaturization of technology right down to the molecular scale [1]. It has now turned into reality in the form of nanotechnology. With most major scientific nations emphasizing on nanotechnology being the driving technology of the new century (through various national nano initiatives, or nano missions), there is considerable progress taking place in this relatively new, yet highly interdisciplinary field of study.

The major interest in nanoparticles has been generated in the research and the industry sector mostly due to its physical and chemical properties which differ quite a lot from the bulk (large scale) materials. At present almost all branches of applied science and technology are being advanced in some way or the other with the advent of nanotechnology. Thus, various new fields such as nanomedicine, nanobiotechnology, nanoelectronics, and so on are emerging in a big way (Fig. 1).

So the question naturally comes to mind what exactly makes the nanomaterials behave differently? The answer can be found by understanding two key concepts of nanotechnology, the enhanced surface to volume ratio and the quantum confinement.

The surface to volume ratio (S/V) is simply the surface area of an object divided by its volume. For symmetrical objects like a spherical particle this ratio is inversely proportional to the radius, and for a cube it is inversely proportional to its sides. What makes a nanomaterial special is that with increasing S/V more and more atoms/molecules of the material become exposed to the surroundings and a larger number of the so-called "dangling bonds" become available at the surface, thus making the particle more active chemically. For a very simple example a spherical particle of ~ 100 nm diameter has only a very small percentage ~ 2 % of its total

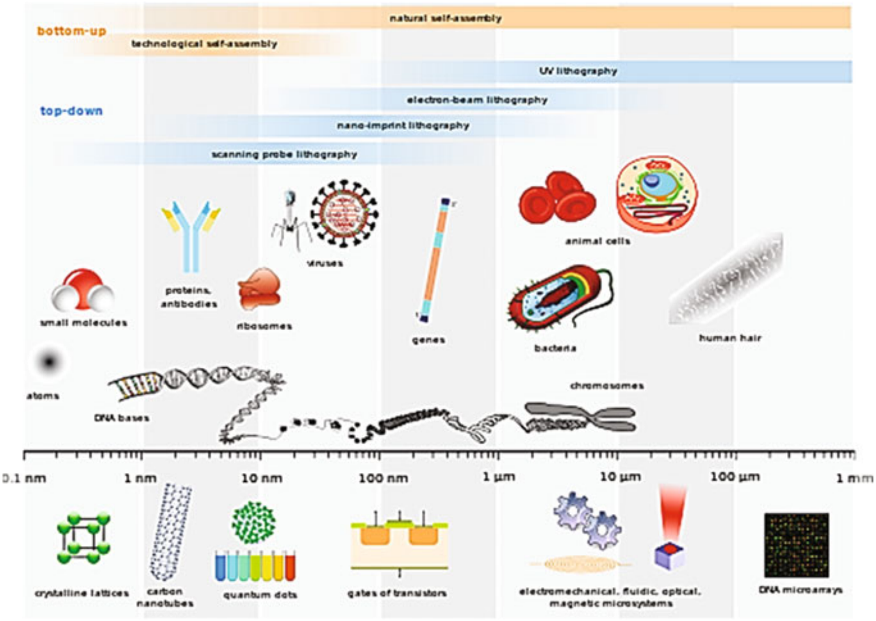


Fig. 1 Comparison of various biological assemblies and technological devices. *Source* Wikimedia Commons, under the Creative Commons Attribution-Share Alike 2.5 Generic license

constituent atoms exposed to the surface, if we go down to particle diameters of ~ 10 nm 20–25 % of the atoms become exposed to the surroundings, while a further size reduction say down to 3 nm increases the fraction of atoms exposed at the particle surface to a 45–60 %. In such circumstances the behavior of the particle becomes altered in terms of its chemical activity. This is precisely the reason behind certain noble metals like Ag, Au, Pt, etc., behaving as highly potent catalysts in their particulate, especially nanoparticulate form. The presence of large number of unsaturated "dangling" bonds on the graphene (a two-dimensional sp^3 bonded honeycomb-like sheet of carbon atoms, which is the building block of graphite) surface which has a very high *S*/*V* makes it an excellent sensing material for chemical and gas detection. Such high *S*/*V* ratio is also responsible for the lowering of melting point in nanocrystals [2–8].

The second vital concept of quantum confinement arises from the size reduction of material leading to the electronic wave functions being more tightly confined and resulting in changes associated with electronic and optical properties of the nanomaterial. A smaller (or bigger) particle size results in a stronger (or weaker) confinement which gives rise to enhancement (or decrease) of the band gap and modifies the band structure of the material. This results in changes in the electron mobility and effective mass, relative dielectric constant, optical properties to name a few. For the reason of quantum confinement effect metallic nanoparticles show interesting properties like variations in color of colloidal suspensions with changing particle size, UV photoemission, enhanced photoluminescence, etc. The light-induced UV emission in Ag nanoparticles has prompted the use of nano-Ag coatings for antimicrobial purposes in the healthcare sector [2–4].

So based chiefly on these aspects and some other aspects of size reduction like phase change, changes in electric/magnetic domains, etc., nanomaterials undergo variations in their melting point, thermal expansion coefficient, hardness, diffusivity, optical absorption, dielectric properties, magnetic properties to name a few. Thus, scientists and engineers today possess a grand ability of significantly alter materials properties by proper size reduction and clever utilization of quantum confinement, which makes nanotechnology so exciting.

2 Some Applications of Nanotechnology

The commercial applications of nanotechnology are truly diverse. Nanotech products are finding their way into a variety of different industries such as automobile, aerospace, biotechnology, cosmetics, defense, energy, electronics, health care, sports/fitness, textiles, footwear. The list below gives few examples how nanotechnology is set to be incorporated in the world around us [2–6].

- Automobile: Tough and ultralight nanocomposite materials for sport cars and high-performance bikes, dirt-resistant paints and anti-fog windshields/mirrors with hydrophobic nanocoatings, better fuel cells for hybrid/electric vehicles.
- Aerospace: Carbon fiber composites for modern aircraft and helicopters that help decrease weight and enhance fuel efficiency. Also the carbon fiber composites help decrease the RADAR profile of combat aircrafts. Shape-shifting aerofoils and winglets in future aircrafts/space shuttles for extreme maneuverability and handling is also a topic of R&D in the aerospace industry.
- Construction: “Self-repairing” concrete made from application of nanocomposites and nanobiotechnology, easy to clean window glass for skyscrapers having coating of nanorods on the surface which can be cleaned only with water without soap/surfactants.
- Defense: Lighter and better nanocomposite helmets and bulletproofing for security/counterterrorism personnel, Self-monitoring, analysis and reporting technology (SMART) uniforms for tactical and special forces are a major research area in the defense sector. SMART weapons and even SMART ammunition may soon be seen in the inventory of the modern militaries of the world.
- Electronics: Nanoscale MOSFETs, carbon nanotube (CNT) FETs, graphene FETs, CNT field emitters, quantum cascade laser, single electron transistors (SET) to name a few.
- Health care: Nano-enabled drug delivery where particular agents are site selectively being delivered to treat complex diseases like cancer, tumors, and

> neurodegenerative diseases like Alzheimer's. The "lab on a chip" or microfluidic testing and analyzing systems to automatically test a large number of pathogens, a more commonplace use is the antibacterial coatings of light activated Nano Titanium in toilets, hospitals, and interiors of public transport systems.

So it is understandable that there has been a nano boom of sorts in recent years. With public and corporate funding flowing for nano-related R&D and manufacturing, it is only making the nanoindustry grow at a considerable pace (Fig. 2).

According to recent reports the global nanotechnology market is estimated to be 2 trillion USD by 2015. Since 1999 there has been an average annual growth rate of about 22 % for nanobiotechnology and about 12–15 % for nanodevices in the world market [5, 11].

It still remains to be seen exactly what form of nanotechnology takes in the coming years, does it challenge old beliefs and keeps coming up with revolutionary changes like the graphene-based devices, CNT FETs, etc., or does it take a more

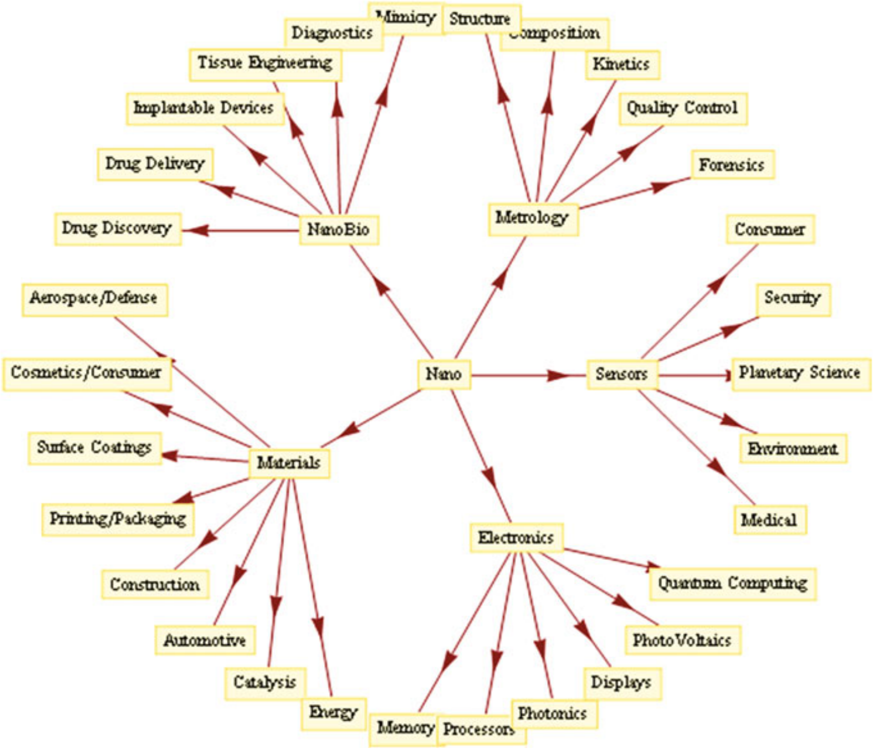


Fig. 2 Schematic of nanotechnology application areas. *Source* WikiMedia Commons, P. Fraundorf, distributed under the Creative Commons Attribution-Share Alike 3.0 Unported license

mellowed evolutionary approach to amalgamate with existing technologies, like in nanomaterials-based composites and nanocrystal-embedded MOS memory devices.

Nevertheless from all the indications over the last few decades it seems that nanotechnology is here to stay and have a significant scientific, technological, and social impact.

In this book we have tried to give a concise overview of the underlying physics and some applications of nanotechnology with a special focus on nanoelectronic devices. The book consists of two sections, in the first section we try to give an overview of the basic physics behind the uniqueness of nanomaterials in terms of the variance of properties from the bulk. We have looked to gradually build up the basic concepts of quantum physics and materials science in order to create understanding of confinement and tunneling effects. Special focus is given to the properties of nanomaterials with size variation. The next section is more application oriented, discussing various nanoelectronic devices such as nanocrystal-embedded MOS memories, spin-torque transfer devices, CNT-based memories, and the nanoscale MOSFET. Special focus has been given on the nanoscale MOSFET as it is a major technology driver for the next generation of electronics.

References

1. http://www.pa.msu.edu/~yang/RFeynman_plentySpace.pdf
2. B. Bhushan (ed.), *Springer Handbook of Nanotechnology* (Springer, Berlin, 2010)
3. H.S. Nalwa (ed.), *Encyclopedia of Nanoscience and Nanotechnology,* vol. 1–10 (2004) and vol. 11–25 (2011) (American Scientific Publishers)
4. H.S. Nalwa (ed.), *Handbook of Nanostructured Materials and Nanotechnology* (Elsevier, Amsterdam, 2000)
5. M.A. Ratner, D. Ratner, *Nanotechnology: A Gentle Introduction to the Next Big Idea* (Pearson, UK, 2002)
6. C.P. Poole Jr, F.J. Owens, *Introduction to Nanotechnology* (Wiley, New York, 2003)
7. W. Farhner (ed.), *Nanotechnology and Nanoelectronics: Materials Devices and Measurement Techniques* (Springer, Berlin, 2005)
8. G.W. Hanson, *Fundamentals of Nanoelectronics* (Prentice Hall, New Jersey, 2007)
9. R. Saito, G. Dresselhaus, M.S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (World Scientific, 1998)
10. S. Datta, *Quantum Transport: From Atom to Transistor* (Cambridge University Press, Cambridge, 2005)
11. Nanotechnology WikiBook (Online) [en.wikibooks.org/wiki/Nanotechnology]

Fundamentals of Quantum Theory

Amretashis Sengupta

> *If you are not confused by quantum physics then you haven't really understood it*
>
> —Niels Bohr

We live in a macroscopic world surrounded by macroscopic objects, which adhere to the laws of classical mechanics and are very comfortable with it. From the trajectory of a humble pebble thrown into a pond and the ripples thus created by it, to the motion of the great planets can all be quite satisfactorily explained within the domain of classical physics.

However, apart from this macroscopic scale, there exist numerous other phenomena around us, especially on the microscopic scale, which seem to defy most conventional "common-sense" beliefs. The key to the understanding of the workings of this amazing microscopic world lies with quantum physics.

According to classical physics and also our everyday understanding a ping-pong ball thrown at a brick wall has absolutely no chance of somehow "magically" going through it and re-appearing on the other side. However in the world of quantum mechanics the almost analogous event of an electron incident upon an energy barrier, does have a finite non-zero probability of appearing on the other side of the barrier (not by some "magic" but rather by a commonplace quantum phenomena called tunneling). What may seem rather absurd (sometimes bordering on ridiculous) in the macroscopic world may be a very common phenomena in the quantum domain.

1 Origins of Quantum Physics

At the beginning of the twentieth century three major issues stood in front of the scientific community that could not be properly explained within the framework of classical physics. First there was the Blackbody radiation problem, which could not

---

A. Sengupta (✉)
School of VLSI Technology, Indian Institute of Engineering Science and Technology,
Shibpur, Howrah 711103, India
e-mail: a.sengupta@vlsi.iiests.ac.in

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_2

be satisfactorily explained by the Rayleigh-Jeans law, especially for high frequencies (the so called "UV catastrophe"). Then there was the photo-electric effect, where the stopping potential for photo-current (electron emission) from an irradiated emitter was found to be dependent on frequency rather than intensity. This was contrary to the classical predictions that for the same frequency higher intensities should produce a higher stopping potential. Thirdly there was the Compton shift, where monochromatic X-rays scattered by a material block always seemed to shift by a certain discrete amount (for a fixed angle of incidence). All these seemed to vaguely point that energy emitted or absorbed in these processes were somehow related to the frequency.

The existence of such fixed amounts of energy or 'quanta of energy' was proposed by Max Planck in his theory of blackbody radiation to successfully explain the radiation spectra. Also in the theory of specific heat of solids, Einstein had considered oscillators vibrating at fixed frequencies to successfully explain the variation of specific heats of solids over a wide range of temperature. Another proof of energy quantization (i.e. having specific discrete levels or 'states' of energy in a system) was given by the success of the Bohr model of the Hydrogen atom. All these theories and experimental evidences were combining to pave the way towards a new theory.

2 The Wave Particle Duality

Between 1924 and 1927 two most significant and very radical ideas emerged, which were to become the cornerstones of what was to emerge as the quantum theory. In 1924 Louis de Broglie, in his thesis titled *Recherches sur la théorie des quanta* proposed the wave-particle duality, a revolutionary concept that hypothized the association of any material particle with an associated matter wave and vice-a-versa.

According to de Broglie hypothesis a material particle travelling with momenta p is associated with a matter-wave of wavelength λ as

$$\lambda = \frac{h}{p} \tag{1}$$

This is also in corroboration with the dual nature of light as observed by various experiments. While classical experiments like the Young's double slit experiment demonstrated the wave nature of light which is responsible for diffraction and interference, the works by Einstein on the photo electric effect is hard evidence of particle nature of light (the particle associated with light energy being known as the photon). In this regard it would be relevant to mention that in 1660–1675 first Pierre Gassendi and thereafter Sir Isaac Newton had hypothized light to be made up of particles (termed as 'corpuscules' by Newton). These ideas had been discarded in favor of the wave theory, which envisioned light to propagate as mechanical waves through a 'luminiferous aether' medium. The existence of the aether was discarded after the famous Michelson-Morely experiment in 1887. With the subsequent discovery of photo-electric effect in 1887 by Heinrich Hertz and its successful theoretical explanation in 1905 by Albert Einstein, the particle theory of light was revived in a whole different avatar and the dual-nature of light emerged.

3 The Uncertainty Principle

In 1927, Werner Heisenberg put forward his uncertainty principle which put rather simply states that there exists a fundamental limit of precision, to which extent specific pairs of complementary variables (also known as canonically conjugate variables) such as position and momenta can be simultaneously measured. To put it mathematically in the form of an equality, let us consider a simultaneous measurement of the position x and the momenta p of a particle. Then Heisenberg's uncertainty principle states that the uncertainty or error in the measurement of position (say Δx) and that of momentum (Δp) must be such that

$$\Delta x \Delta p \geq h \tag{2}$$

This error in measurement is a fundamental limit in nature itself and cannot be improved upon.

These two basic postulates of quantum mechanics can be correlated in a very interesting way. Let us consider a microscopic particle moving around freely in space and an observer equipped with a powerful microscope who is looking for it. Now we consider energetic photons colliding with the target particle and being reflected back into the microscope thus giving out the position of the particle. Now if a sole photon of wavelength λ were to collide with the particle, then according to de Broglie hypothesis the change in the momentum of the particle should be of the order of the photon momentum which is $p = h/\lambda$. Also considering the wavelength of the photon, the uncertainty in position should also be $\geq \lambda$. Which implies the product of the two uncertainties should be $\geq h$, which is essentially Heisenberg's uncertainty principle. Now if we look to minimize our error in measurement of position by minimizing lambda (say the hypothetical gamma-ray microscope) we would measure position much more accurately but in the process sacrifice accuracy in measuring the momenta.

4 The Wave Packet and the Wave Function Ψ

Now if we look to associate microscopic particles with a matter wave then we are to require a mathematical expression or simply a wave function which we call Ψ. This function is required to have both spatial and temporal dependence and therefore

$$\Psi(\vec{r},t) = \psi(\vec{r})\exp(-i\omega \cdot t) \tag{3}$$

For the 1-dimensional case this reduces to

$$\Psi(x,t) = \psi(x)\exp(-i\omega \cdot t) \tag{4}$$

With this wave function Ψ, we represent the matter wave associated with the microscopic particle under consideration. Now consider a wave group or a "wave packet" associated with a particle, which is this wave function Ψ confined in a small region of space. As a wave packet could be considered to be made by the superposition of a number of plane waves with infinitesimally differing k-values (wave vector) we could consider

$$\Psi(\vec{r},t) = \sum_{n=1}^{\infty} c_n(k)\Psi_n(\vec{r},t) \tag{5}$$

where $c_n(k)$ is the corresponding normalizing constant.

Now if we are to know the exact position of the particle within that wave packet then a narrower wave group (i.e. a smaller λ) would suit us better. But if λ becomes small, then the error in determining the momenta increases as $p = h/\lambda$ according to de Broglie hypothesis. Thus the concept of wave packets is consistent with Heisenberg's uncertainty principle as well.

As for the position of the particle, it may be considered that the particle is most likely to be at the center of the wave packet, but the probability of its existence in other regions of the wave packet are essentially non-zero. The probability of finding the particle at any particular position x is given by the quantity $\Psi^*\Psi$. Also since the particle does indeed exist, thus we could have the condition

$$\int_{-\infty}^{+\infty} \Psi^*\Psi \mathrm{d}\tau = 1 \tag{6}$$

which is famously known as the normalization condition. This essentially means that the probability of finding the particle over $-\infty$ to $+\infty$ is essentially 1. $\mathrm{d}\tau$ is a volume ensemble.

Another important quantity known as the probability current density is given by

$$J(\vec{r},t) = \frac{i\hbar}{2m}(\Psi\nabla\Psi^* - \Psi^*\nabla\Psi) \tag{7}$$

The significance of this quantity is the particle number conservation, (7) essentially means that if the probability of finding the particle increases in a specific region of space, then in other regions this probability must decrease.

The function Ψ should also fulfill the following conditions

1. It must be a finite and single valued function. This is essential because if the wave function is infinite then the probability of the finding the particle also becomes infinite which is impossible. Also the single valued-ness is essential so as to eliminate the un-physical concept that the particle has two different probabilities of existing at the same point.
2. The derivative of Ψ should be single valued and continuous. This is quite like the boundary conditions in electromagnetics where the derivative of the potential is considered continuous to properly account for interfaces and boundaries of different media. This is also true in most cases of quantum mechanical potential berries and finite wells. Even in the case of a particle confined within an infinite potential well there exists a small trailing edge of the wavefunction beyond the boundaries of the well.
3. $\Psi \to 0$ when $r \to \pm\infty$, which ensures $\int |\Psi|^2 \mathrm{d}\tau$ to be finite over all space.

With the rules of the game now laid down, we proceed to the equation that would describe the dynamics of the microscopic particle in the quantum realm, which is in fact the famous Schrödinger equation.

5 The Schrödinger Equation

It would be useful if we just recollect a little bit of operator algebra. Say we have a function $f(x)$ and an operator $\widehat{\Theta}$, and the operation be expressed as

$$\widehat{\Theta} f(x) = \alpha f(x) \tag{8}$$

then $f(x)$ is known as the eigenfunction of the operator $\widehat{\Theta}$, and α the eigenvalue of the same. In case of quantum mechanics, the measurement of any dynamical quantity g_n can be made from the wave function by means of an operation by a suitable operator $\widehat{\Gamma}$ corresponding to that quantity g_n.

$$\widehat{\Gamma}\Psi_n = g_n \Psi_n \tag{9}$$

Ψ_n is known as the eigenstate of the system and g_n the corresponding eigen-value. Now g_n can be various dynamical variable, depending on the operator $\widehat{\Gamma}$. We could measure quantities like position, momenta, energy, angular momenta etc. of the system by choosing suitable operators, as we shall soon see.

Another important thing that should be noted is that, as soon as we make a measurement on the system and find a particular eigenstate, the whole system collapses to that particular eigenstate. Any further measurements also give the same eigenvalue for the given operator.

Now in order to have a proper wave equation, we expect it to be a linear differential equation in $\Psi(\vec{r}, t)$, as the wave function must be linear to preserve the superposition of waves. The equation should not also contain any dynamical variable so that we may continue to get wave functions of the packet corresponding to different dynamical variables. Also as a requirement, for a free particle the solution of the said equation should correspond to infinite plane waves. This is represented by

$$\Psi(\vec{r}, t) = A(k)\exp\left\{i\left(\vec{k}\cdot\vec{r} - \omega t\right)\right\} \tag{10}$$

From this plane wave function we can find out the energy and momentum operators easily. For this we use the wave vector k may be expressed as $k = p/\hbar$. Also from Einstein's theory of photo-electric emission the energy quanta for a radiation of frequency is given by $E = h\nu$ which gives, $\omega = \frac{E}{\hbar}$. Thus differentiating (10) w.r.t. time we find

$$\begin{aligned}\frac{\partial\Psi(x,t)}{\partial t} &= -\frac{i}{\hbar}E\cdot A\exp\left\{\frac{i}{\hbar}(px - Et)\right\}\\ &= -\frac{i}{\hbar}E\cdot\Psi(x,t)\end{aligned} \tag{11}$$

Equation (10) gives us the energy operator

$$\widehat{E} = i\hbar\frac{\partial}{\partial t} \tag{12}$$

With this operator the energy eigenvalue of a corresponding energy eigenstate can be found out. Similarly it is possible to find out the momentum operator form (10) by differentiating w.r.t. x as

$$\widehat{p} = -i\hbar\frac{\partial}{\partial x} \tag{13}$$

By taking the second derivative of (10) w.r.t. x, and rearranging and dividing by $2m$, we arrive at

$$\begin{aligned}-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} &= \frac{p^2}{2m}\Psi(x,t)\\ &= E\Psi(x,t)\end{aligned} \tag{14}$$

thereafter using (11) we arrive at

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} = i\hbar\frac{\partial\Psi(x,t)}{\partial t} \tag{15}$$

In 3-dimensions this could be extended to

$$\nabla^2\Psi(\vec{r},t) + \frac{2mE}{\hbar^2}\Psi(\vec{r},t) = 0 \tag{16}$$

this is the simplest form of the time dependent Schrödinger equation. If there exists a force field in which the particle is moving then the Schrödinger equation is modified to a more general form. The total potential associated with the field $V(x)$ adds to the energy $E = \frac{p^2}{2m}$ in this case and thus

$$E = \frac{p^2}{2m} + V(\vec{r},t) \tag{17}$$

Which combined with (16) leads to

$$i\hbar\frac{\partial\Psi(\vec{r},t)}{\partial t} = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r},t)\right]\Psi(\vec{r},t) \tag{18}$$

From (18), we may define an operator which is known as the Hamiltonian operator or simply the Hamiltonian $\hat{H}$ of the system given by

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r},t) \tag{19}$$

The general form of (18) is written as

$$\nabla^2\Psi(\vec{r},t) + \frac{2m}{\hbar^2}[E - V(\vec{r},t)]\Psi(\vec{r},t) = 0 \tag{20}$$

The Eq. (20) is known as the time dependent Schrödinger equation. In many cases the Hamiltonian may not explicitly depend on time, and for such cases V is a function of position only. In such cases the time independent form of the Schrödinger equation is used, which is

$$\nabla^2\psi(\vec{r}) + \frac{2m}{\hbar^2}[E - V(\vec{r})]\psi(\vec{r}) = 0 \tag{21}$$

In $\psi(\vec{r})$ is the position dependent part of the wave function

$$\Psi(\vec{r},t) = A(k)\psi(\vec{r})\exp\left(-\frac{iEt}{\hbar}\right) \tag{22}$$

6 The Expectation Value

The solution of Schrödinger equation gives us results for various parameters of the system within the limits of the uncertainty principle. The quantities that result from a measurement performed are probabilities rather than exact values. Thus we consider the expectation value of a certain parameter, say position, which is like the average value of position obtained for a large number of particles described by the same wave function at a certain instant of time.

The expectation value of a certain quantity g associated with an operator $\hat{\Gamma}$ is defined as

$$\langle g \rangle = \int_{-\infty}^{+\infty} \Psi^* \hat{\Gamma} \Psi \mathrm{d}\tau \tag{23}$$

This indicates that at a certain instant of time t if we make a substantially large number of measurements over the wavefunction in order to extract the value of the quantity g (which could be position, momentum, energy etc.) we are most likely to arrive at this particular value of g, which is thus known as the expectation value.

Thus depending upon the quantity for which the measurement is made, the expectation value is determined with the suitable operator. Let us take the case of position for example. Let the probability of finding a particle at a position x_j in the interval dx be given by P_j, thus

$$P_j = \left|\Psi(x_j)\right|^2 \mathrm{d}x \tag{24}$$

For a large number measurements, the expectation value would be

$$\langle x \rangle = \frac{\int_{-\infty}^{+\infty} x|\Psi|^2 \mathrm{d}x}{\int_{-\infty}^{+\infty} |\Psi|^2 \mathrm{d}x} \tag{25}$$

If Ψ is a normalized wave function then the denominator becomes unity and we have the expectation value for position

$$\langle x \rangle = \int_{-\infty}^{+\infty} x|\Psi|^2 \mathrm{d}x \tag{26}$$

Which can be written in as

$$\langle x \rangle = \int_{-\infty}^{+\infty} \Psi^* \hat{x} \Psi \mathrm{d}x \tag{27}$$

where $\hat{x}$ is the position operator defined as $\hat{x}\Psi = x\Psi$.

7 The Free Particle Solution

The free particle such as an electron moving in free space can be described with the Schrödinger equation (for simplicity let's assume only the x component) as

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + \frac{2mE}{\hbar^2}\psi = 0 \tag{28}$$

the solution for which is given by

$$\psi(x) = A\mathrm{e}^{ikx} + B\mathrm{e}^{-ikx} \tag{29}$$

where, $k = \sqrt{\frac{2mE}{\hbar^2}}$. The energy solution for this is given by

$$E = \frac{\hbar^2 k^2}{2m} \tag{30}$$

Which is essentially a parabola. When such an electron is confined within a region of space it can be modeled (ideally) as a particle within a rigid box (an infinite quantum well). Such a quantum well of length L in x direction can be represented by a potential

$$\begin{aligned} V(x) &= 0 \quad \text{for } 0 \le x \le L \\ V(x) &= \infty \quad \text{otherwise} \end{aligned} \tag{31}$$

In such cases, the solution of the Schrödinger equation gives quantized energy levels (as discussed in detail in the next chapter). It can also be elegantly deduced using the basic wave particle duality and the concept of standing matter waves (Fig. 1).

From the wave particle duality $\lambda = h/p$. For the matter waves within the well the situation is analogous to a standing waves in a string whose both ends are tied. For the classical tied string concept we know

$$n\lambda = 2L \tag{32}$$

n represents the mode of the vibration $(n = 1, 2, 3\ldots)$. Further if we consider the particle nature we know kinetic energy $E_\mathrm{K} = \frac{p^2}{2m}$. Which is the total energy in this case as inside the well $V(x) = 0$. Thus in terms of de Broglie wavelength of the particle we may write the energy of the

$$E_n = \frac{h^2}{2m\lambda^2} = \frac{n^2h^2}{8mL^2} \tag{33}$$

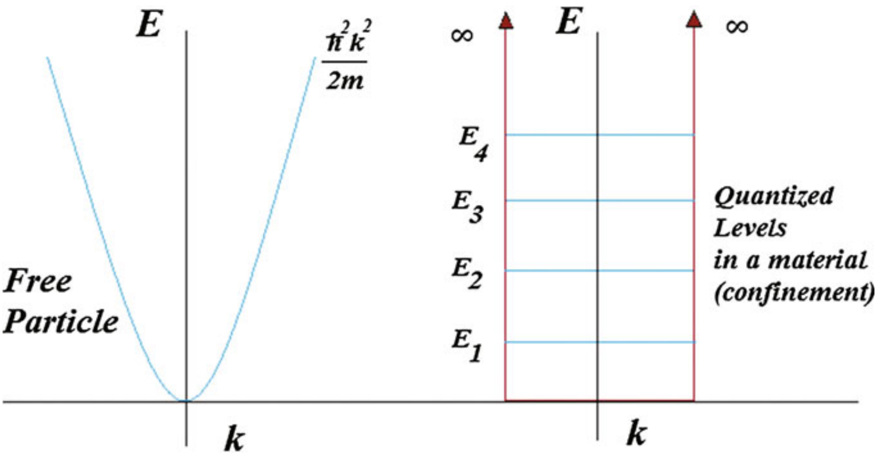


Fig. 1 The continuous energy solution for a free particle (such as electron in free space) becomes discrete energy levels for a confined system (such as an electron inside an infinite quantum well)

for $n = 1, 2, 3\ldots$. So we see how confining the electron within a rigid box can lead to the so-called energy quantization, which is a very fundamental concept in quantum mechanics. A more rigorous fully quantum mechanical proof by solution of the Schrödinger equation with boundary conditions for this particle in a box problem is provided in the next chapter.

8 The Linear Harmonic Oscillator Problem

Another one of the simplest problems in quantum physics that demonstrates key concepts such as energy quantization is the treatment of the linear harmonic oscillator. Physically such an oscillator can be thought of as a particle undergoing harmonic oscillations for small amplitudes in one dimension (say x). This can be described in line with the classical case of a particle under the influence of a parabolic potential represented by

$$V(x) = \frac{1}{2}kx^2 \tag{34}$$

k being the so-called spring constant. Classically such an oscillator would have frequency (Fig. 2)

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{35}$$

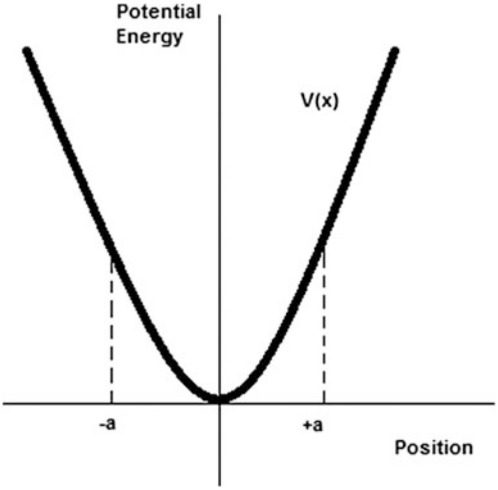


Fig. 2 Potential energy of a linear harmonic oscillator

For a quantum mechanical system under the influence of such a potential, the Schrödinger equation would be of the form

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}\left(E - \frac{1}{2}kx^2\right)\psi = 0 \tag{36}$$

Equation (30) can be simplified by introducing the quantities λ and α given by

$$\lambda = \frac{2mE}{\hbar^2} \tag{37}$$

$$\alpha = \frac{2\pi mv}{\hbar} \tag{38}$$

Thus the Eq. (30) becomes

$$\frac{d^2\psi}{dx^2} + \left(\lambda - \alpha^2 x^2\right)\psi = 0 \tag{39}$$

For large values of x, i.e. $\alpha^2 x^2 \gg \lambda$, (33) reduces to

$$\frac{d^2\psi}{dx^2} - \alpha^2 x^2 \psi = 0 \tag{40}$$

The general solution for which is

$$\psi(x) = e^{\pm\frac{\alpha}{2}x^2} \tag{41}$$

as the solution $\psi(x) = \mathrm{e}^{+\frac{\alpha}{2}x^2}$ blows up as $x \to \infty$, hence we take only the other solution

$$\psi(x) = \mathrm{e}^{-\frac{\alpha}{2}x^2} \tag{42}$$

However in cases the value of x is small and the value of λ can no longer be neglected. In such cases, we can solve the problem using a power-series expansion of an introduced function $f(x)$ to express $\psi(x)$ as

$$\psi(x) = \mathrm{e}^{-\frac{\alpha}{2}x^2} \cdot f(x) \tag{43}$$

Putting this value of $\psi(x)$ in (33) and rearranging, we arrive at

$$f'' - 2f'\alpha x + (\lambda - \alpha)f = 0 \tag{44}$$

Defining a variable $\rho = \sqrt{\alpha}x$, and rearranging, (38) can be recast as

$$\frac{\mathrm{d}^2H}{\mathrm{d}\rho^2} - 2\rho\frac{\mathrm{d}H}{\mathrm{d}\rho} + \left(\frac{\lambda}{\alpha} - 1\right)H = 0 \tag{45}$$

where $H(\rho)$ is the Hermite's polynomial and thus ψ is expressed as

$$\psi(\rho) = \mathrm{e}^{\frac{-\rho^2}{2}}H(\rho) \tag{46}$$

The Hermite polynomial can be expressed as

$$H(\rho) = \sum_n a_n\rho^n \tag{47}$$

$n = 0, 1, 2, \ldots H(\rho)$ follows the recursion relation,

$$\frac{a_{n+2}}{a_n} = -\frac{\left(\frac{\lambda}{\alpha} - 2n - 1\right)}{(n+1)(n+2)} \tag{48}$$

From the general solution in (40), taking only the cases where the solution satisfies the requirement that $\psi(\rho) \to 0$ as $\rho \to \infty$, we arrive at the following condition using (40)–(42)

$$\frac{\lambda}{\alpha} = (2n + 1) \tag{49}$$

Since $\frac{\lambda}{\alpha} = \frac{2E}{h\nu}$, thus we have

$$\begin{aligned} E_n &= \left(n + \frac{1}{2}\right)h\nu \\ &= \left(n + \frac{1}{2}\right)\hbar\omega \end{aligned} \tag{50}$$

where n is 0, 1, 2, 3…. This is indeed an interesting solution, which implies that only certain energy values are allowed for a quantum linear harmonic oscillator. This revelation that in the quantum domain, the energy that a system may possess is not continuous but rather discretized lies at the heart of quantum physics.

For the case $n = 0$ the system possesses a non-zero energy given as

$$E = \frac{1}{2}\hbar\omega \tag{51}$$

This is the famous "zero-point energy". This concept is unique to quantum physics as classically the lowest possible energy of a harmonic oscillator should be zero. This indicates that when the system is at its lowest possible energy state (i.e. temperature nearing absolute zero) it is still supposed to have a non-zero energy value (Fig. 3).

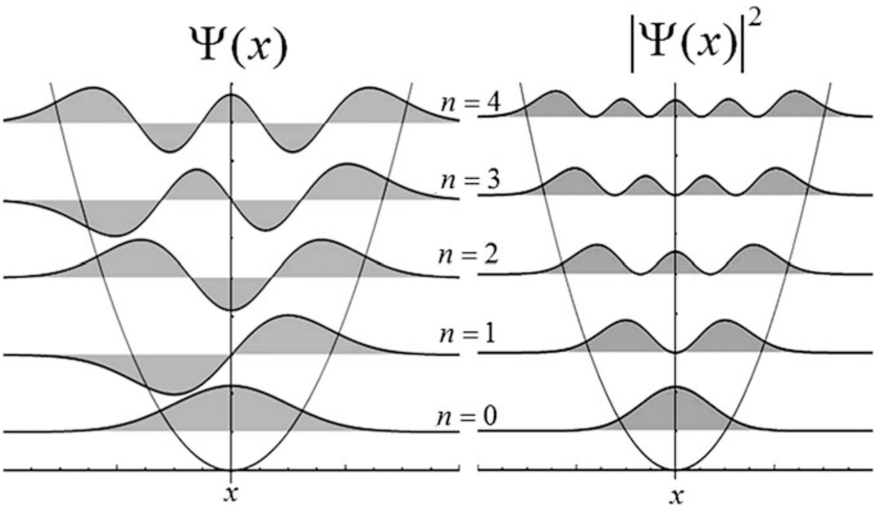


Fig. 3 Nature of wavefunctions and corresponding probabilities for the harmonic oscillator problem (Courtesy Wikimedia Commons. This is an edited version: original file © Alan McC, distributed under the creative commons license)

The wave functions for the different states are expressed as

$$\psi_n(\rho) = \left(\frac{2mv}{\hbar}\right)^{1/4} (2^n n!)^{-1/2} H_n(\rho) \mathrm{e}^{-\rho^2/2} \quad (52)$$

Interestingly there exists a small non-zero part of the wavefunction beyond the potential barrier. This tail of the wavefunction outside the bound system represents a probability that a particle may have a non-zero probability of existing beyond the confining potential, which is the basic concept of quantum mechanical tunneling.

9 The Kronig-Penney Model for Electron in a 1-Dimensional Lattice

In case of an electron confined in a rigid box since $V(x) = 0$ everywhere inside the box, it can be considered as an “empty solid” (i.e. just a confining potential without any atoms or molecules within it). In a real solid however atoms and molecules exist and the potential inside the boundaries is not zero but rather periodic according to the lattice (the concept of lattice and structure of solids are discussed in detail the subsequent chapters, for now we can assume lattice as a neatly arranged grid/blueprint where atoms/molecules are located in a solid material) (Fig. 4).

The simplest solid to consider would be a linear chain of equally spaced atoms. In such a linear chain the potential would look something as shown in the figure. Considering an infinite chain, this was approximated by a periodic square well potential by Kronig and Penney.

In such a periodic potential, where

$$U(x) = U(x + L) \quad (53)$$

The Bloch’s theorem expresses the wavefunction with a periodic amplitude part $u(x)$ having the same periodicity as that of the potential (1-D lattice in this case). The wave function is represented as

$$\psi(x) = \mathrm{e}^{i\vec{k}\cdot\vec{x}} u(x) \quad (54)$$

The amplitude part satisfies the condition $u(x) = u(x + L)$, which ensures the continuity of the wavefunction in the periodic lattice (Fig. 5). This also means that a translation by the periodicity L, results in only a phase shift of the original wave by an amount $\mathrm{e}^{i\vec{k}\cdot\vec{L}}$. Mathematically this can be written as

$$\psi(x + L) = \psi(x) \mathrm{e}^{i\vec{k}\cdot\vec{L}} \quad (55)$$

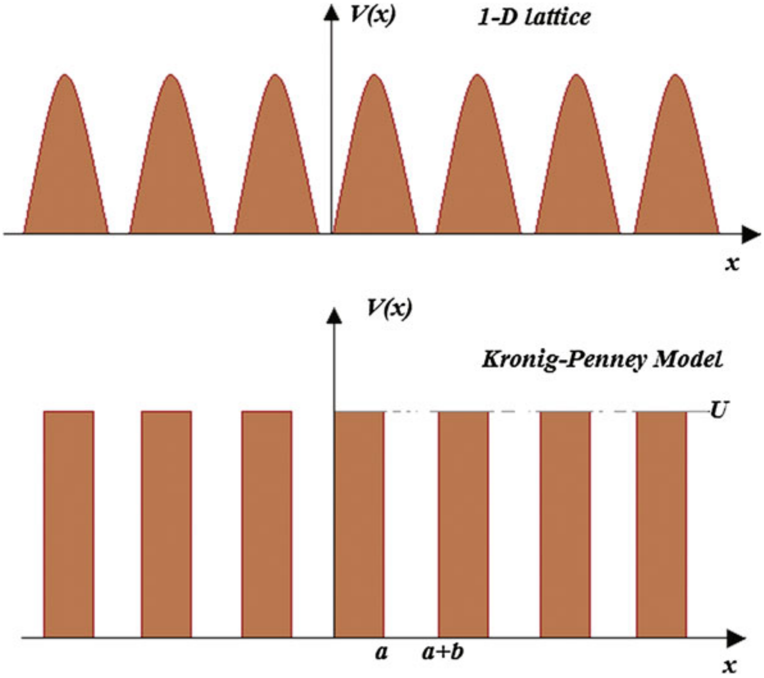
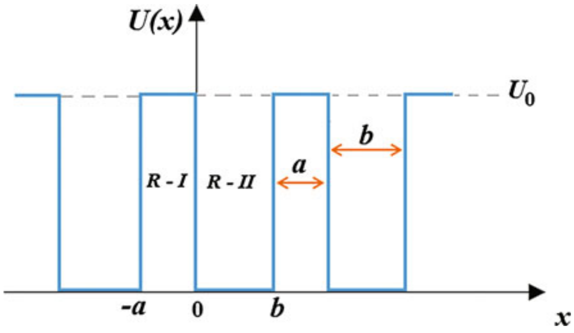


Fig. 4 Potential profile of 1-D lattice

We now direct our attention to a toy example of a 1-D linear chain of atoms. Let us consider the diameter of the positively charged ion-cores to be a and the inter-atomic spacing to be b. The lattice parameter in such a case can be written as $L = (a + b)$. For such a lattice the potential has two distinct regions being repeated along the x direction. The repeat unit being within $0 \leq x \leq (a + b)$. If we solve the Schrödinger equation for this repeat unit consistent with the Bloch's theorem, it is possible to have the solution for the whole 1-D lattice.

Fig. 5 A typical periodic potential in 1-D lattice



In region I (R-I i.e. $0 \le x \le a$) the potential is $U(x) = U_0$ which is constant. Thus in R-I the Schrödinger equation is given by

$$\frac{\mathrm{d}^2\psi_\mathrm{I}}{\mathrm{d}x^2} + \frac{2m}{\hbar^2}(E - U_0)\psi_\mathrm{I} = 0 \tag{56}$$

This has the general solution

$$\psi_\mathrm{I} = A\mathrm{e}^{\beta x} + B\mathrm{e}^{-\beta x} \tag{57}$$

In (56) $\beta = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}}$. The solution for the region-II (R-II) where the potential $U(x) = 0$ is given by

$$\psi_\mathrm{II} = C\mathrm{e}^{i\alpha x} + D\mathrm{e}^{-i\alpha x} \tag{58}$$

In (57) $\alpha = \sqrt{\frac{2mE}{\hbar^2}}$. Using the boundary conditions

$$\begin{aligned} \psi_\mathrm{I}|_{x=0} &= \psi_\mathrm{II}|_{x=0} \\ \left.\frac{d\psi_\mathrm{I}}{dx}\right|_{x=0} &= \left.\frac{d\psi_\mathrm{II}}{dx}\right|_{x=0} \end{aligned} \tag{59}$$

We arrive at a the relations

$$\begin{aligned} A + B &= C + D \\ \beta(A - B) &= \alpha(C - D) \end{aligned} \tag{60}$$

Further as the solution in a periodic potential should be a Bloch wave, we may write

$$\psi_\mathrm{I}(-a) = \psi_\mathrm{II}(b)\mathrm{e}^{ikL} \tag{61}$$

Using (57)–(61), we arrive at the equality

$$\cos(kL) = \frac{\beta^2 + \alpha^2}{2\alpha\beta}\sinh(\alpha a)\sin(\beta b) + \cosh(\alpha a)\cos(\beta b) \tag{62}$$

By taking the limit $b \to 0$ and $U_0 \to \infty$ (which mathematically signifies a dirac delta type periodic potential) and taking the constant $P = \frac{m_\mathrm{e} abU_0}{\hbar^2}$ (62) can be simplified to

$$\cos(ka) = \frac{P\sin(\alpha a)}{\alpha a} + \cos(\alpha a) \tag{63}$$

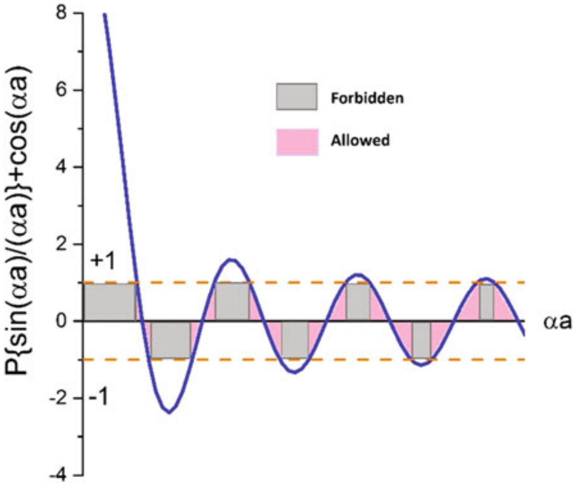


Fig. 6 The graphical solution of (63) showing the allowed and the forbidden regions of the solution

The solution is a transcendental one, and it needs to be solved graphically. A graphical solution involves the plotting of the RHS and the LHS of (63) and finding the regions where solutions are possible.

The graphical solution shows that the relation in (63) is satisfied for only certain ranges of values of αa for which the value of the RHS is within +1 and −1 (i.e. the maximum and the minimum value of the cosine function on the LHS, indicated by the orange dashed lines) (Fig. 6). The regions in which the RHS value exceeds this limit, the corresponding values of αa are not allowed and these forbidden regions are highlighted in grey. Other regions having allowed values are highlighted in purpule.

As $\alpha = \sqrt{\frac{2mE}{\hbar^2}}$, the significance of the solution of (63) means that for the 1-D lattice the energy E can't have continuous values as in the case of a free particle (i.e. the relation $E = \frac{\hbar^2 k^2}{2m_e}$ is not allowed for all values of the wave vector k). For certain values of k there exists a discontinuity in the E–k plot (Fig. 7).

Physically this existence of discontinuity in the energy band structure signifies the forbidden-gap in the bandstructure of a 1-D solid.

The points of discontinuity are $k = \pm\frac{n\pi}{a}$, where $n = 1, 2, 3\ldots$. These regions in the k-space for which there exists allowed values of E (e.g. within $k = \frac{n\pi}{a}$ to $-\frac{n\pi}{a}$) are known as the Brillouin Zones of the lattice. Physically the electron travelling in the lattice can be considered to be reflected back from the boundaries of the subsequent Brillouin Zones. A more detailed description of these zones are provided in the chapter on solid state physics.

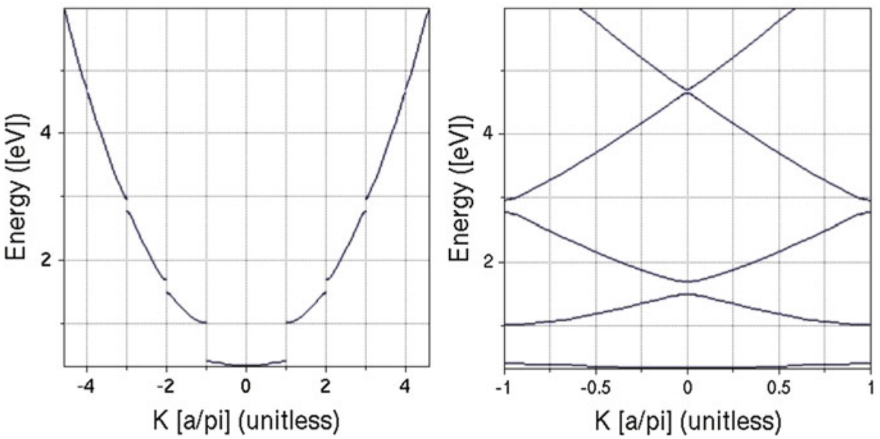


Fig. 7 *E*–*k* relation of 1-D periodic lattice plotted in **a** the extended and **b** the reduced zone scheme

This can be plotted in the extended zone scheme or the reduced zone scheme (i.e. by applying Bloch's theorem the *E*–*k* relation in regions outside $k = \pm\frac{\pi}{a}$, are "folded back" into the same by shifting the phase accordingly).

Thus the simple Kronig-Penney model gives an insight into the quantum mechanical origin of energy band gaps in a 1-D solid. In case of bulk (3-D) solids the potential is not as simple to visualize and involves electro-electron interactions as well. An accurate description of such materials can be obtained by tight binding or ab initio methods. However as a starting point in quantum physics of solids the Kronig-Penney model retains a special significance.

10 Chapter Summary

Here we have presented in brief the key concepts of quantum physics such as the uncertainty principle, wave particle duality, the basic laws that govern matter waves and seen how to set-up the Schrödinger equation for a quantum mechanical system. Quantum phenomenon such as energy quantization through confinement and origin of band gaps in solids has also been introduced. In the subsequent chapter we will explore more about the tunneling and other phenomena in quantum physics. With these very basic concepts of quantum physics we would progress further through the book to develop our understanding of nanomaterials and their physical properties and some novel nanoscale devices.

Further Reading

1. R. Eisberg, R. Resnick, *Quantum Physics of Atoms Molecules, Solids, Nuclei and Particles* (Wiley, London, 1985)
2. A. Beiser, *Perspectives of modern Physics* (McGraw Hill, New York, 1968)
3. A. Ghatak, *Basic Quantum Mechanics* (McMillan, London, 2002)
4. D.J. Griffiths, *Introduction to Quantum Mechanics*, 2nd edn. (Prentice Hall, New Jersey, 2004)
5. J.J. Sakurai, S.F. Tuan, *Modern Quantum Mechanics*: *Revised Edition*. (Pearson Education, New York City, 1994)
6. B.H. Bransden, C.J. Joachain, *Quantum Mechanics* (Pearson, New Jersey, 2000)
7. G. Aruldhas, *Quantum Mechanics* (Prentice Hall, New Jersey, 2008)
8. G.W. Hanson, *Fundamentals of Nanoelectronics* (Prentice Hall, New Jersey, 2007)
9. C.P. Poole Jr, F.J. Owens, *Introduction to Nanotechnology* (Wiley Interscience, NY, 2003)
10. W. Farhner (ed.), *Nanotechnology and Nanoelectronics: Materials Devices and Measurement Techniques* (Springer, Berlin, 2005)
11. B. Bhushan (ed.), *Springer Handbook of Nanotechnology* (Springer, Berlin, 2010)

Basic Solid-State Physics and Crystallography

Amretashis Sengupta

The physics and structure of solid materials is an extremely vast and interesting field of study. In this chapter, we would only look at the basic concepts of solid-state physics, with an outlook toward nanocrystalline solids.

The solids can be briefly classified into two kinds: the crystalline solids and the amorphous solid materials, while most solids in nature are crystalline having their atoms/molecules arranged in a regular repetitive pattern in the three dimensions. This atomic scale blueprint of a material, which but a set of infinite, arranged points related to each other by certain symmetry operations (translation, reflection, rotation etc.) is known as the crystal lattice (Fig. 1).

The lattice itself is not a physical entity rather a geometrical structure of points in space having some definite arrangement (depending upon the type of crystal). When each and every lattice point is associated with some set of atoms/molecules (which are very much physical objects) it forms the crystal structure. The lattice structure is material dependent and is governed by the type of bonds that lead to crystal formation.

The amorphous material is a system where no specific arrangement of atoms/molecules exists in long range. The structure of an amorphous solid may be envisioned as a snapshot of a liquid whose molecules are immobile owing to very high viscosity (viz. the super-cooled liquid phase). There may be very short range order of arrangement (extending to a few of atoms/molecules) but however the periodicity does not extend throughout the material.

The crystalline solids are held together by various types of bonds depending upon the size and electronegativity of the constituent elements. Among the bonds, there exist four types.

A. Sengupta (✉)
School of VLSI Technology, Indian Institute of Engineering Science and Technology,
Shibpur, Howrah 711103, India
e-mail: a.sengupta@vlsi.iiests.ac.in

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_3

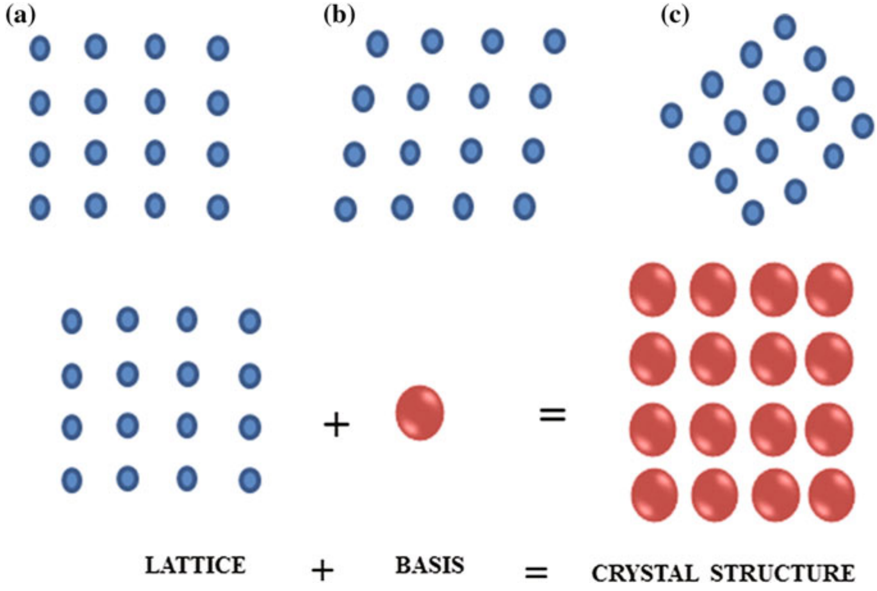
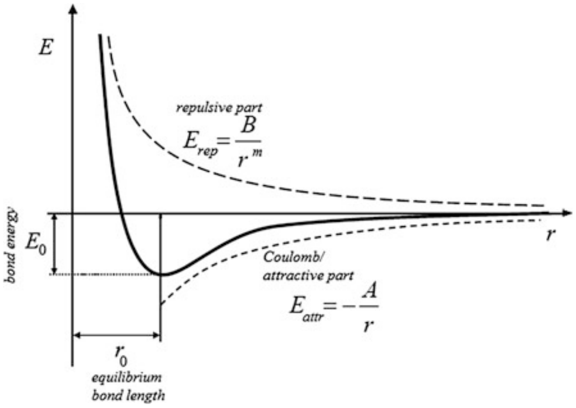


Fig. 1 Concept of lattice and basis

1 Ionic Bonds

These bonds occur in solids made up of atoms differing by a large amount in terms of electronegativity. In such cases to attain a stable configuration, electron transfer takes place among the constituent atoms and anion/cation pairs are formed. In such ionic solids (like NaCl, CsClZnS, CaF_2 etc.), the bonding potential energy consists of two parts, one is the Coulomb part which is attractive in nature, the other is a repulsive part which has its origins in the Pauli's exclusion principle. As atoms look

Fig. 2 Ionic bonding potential energy profile



to come too close to each other, the electron clouds look to repel each other so as not to have two electrons in the same state which is the origin of the repulsive force in this case (Fig. 2).

The Coulomb potential is of the form $-\frac{A}{r}$ while the repulsive part has the form $\frac{B}{r^m}$. The value of A is $\frac{Z_1Z_2e^2}{4\pi\varepsilon_0}$ while B has to be determined empirically. The value of m is usually 8. The two forces create an interesting resultant potential profile. The minima of the resultant potential represent the equilibrium bond length r_0 and the bond energy E_0.

2 Covalent Bonds

The covalent bond is formed due to the sharing of electrons between atoms having same or comparable electronegativity. The sharing takes place in order to achieve filled or half-filled molecular orbitals, which gives the constituent atoms a more stable configuration (Fig. 3).

The physical description of the sharing of electrons can be considered as electrons tunneling back and forth between the potential wells due to the core of the individual atoms. The frequency of such tunneling is usually very high with transit times of the order of 10^{-14} s.

In such bonds depending upon the overlap of the atomic orbitals various types of bonds can be formed, if the overlap is between one s orbital and one p orbital such overlap is known as sp bonding. If it involves one s and two p orbitals, it is sp2 and similarly one s and three p orbitals overlap to produce sp3 bonds. End-to-end overlap of orbitals gives rise to the σ bonds while side-wise (lateral) overlap creates π bonds.

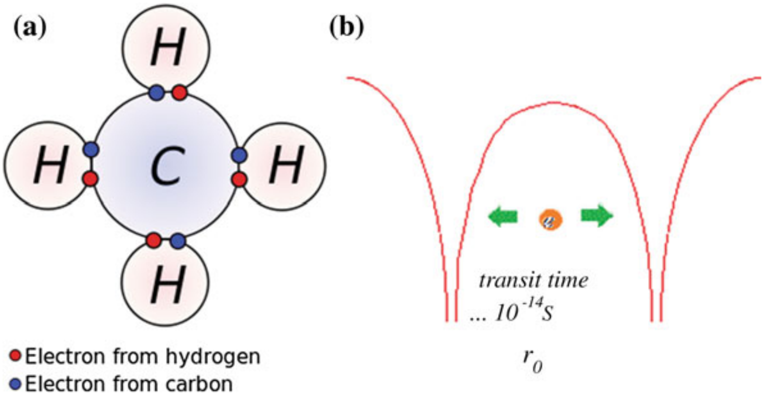


Fig. 3 **a** Covalent bond in a methane molecule CH_4 due to sharing of electrons and **b** schematic of the sharing of electrons from quantum mechanical perspective

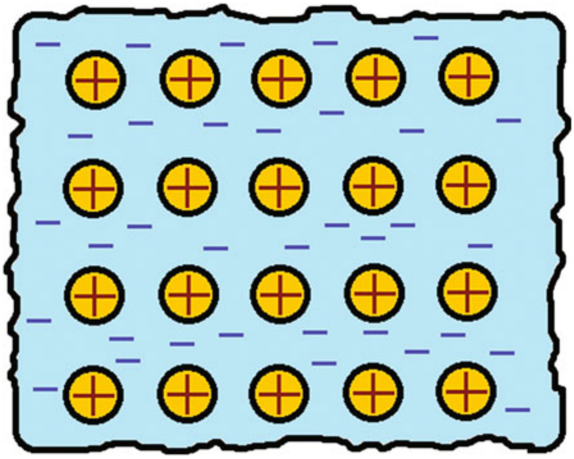


Fig. 4 Schematic of metallic bonds between positive ion cores and free electron gas

3 Metallic Bonds

In metals there exist a large number of so-called free electrons which may be envisaged as a gas of electrons. The atom cores are, however, localized at the designated lattice points. Metallic bonds are thus conceptualized as bonding resultant from the coulomb attraction between these positively charged cores and the electron gas that permeates the metallic mass (Fig. 4).

4 Van der Waals Bonds

These are the secondary bonds having much lesser bonding energies compared to the ionic, covalent, or metallic bonds and mostly originate because of dipole–dipole interactions in atoms/molecules. If they occur between polar molecules (i.e., molecules with permanent dipole moment such as water, HCl etc.) they are termed as dipole–dipole interactions. Van der Waals bonds also include the hydrogen bonding in this category.

If one of the participating molecules is polar and the other does not have a permanent dipole moment (i.e., nonpolar) then a dipole moment may be induced in the nonpolar molecule. The secondary bonds formed by this mechanism fall under the category of Van der Waals bonding as well.

Finally the fluctuating dipoles, due to the asymmetry of the electron cloud in molecules can also lead to Van der Waals bonds.

Usually, all materials exhibit contributions from one or more type of these four kinds of bonding depending upon the constituent atoms/molecules. It is the percentage of contribution that a certain kind of primary bond makes to a solid is what makes it termed as ionic bonded, covalent bonded, or metallic solid.

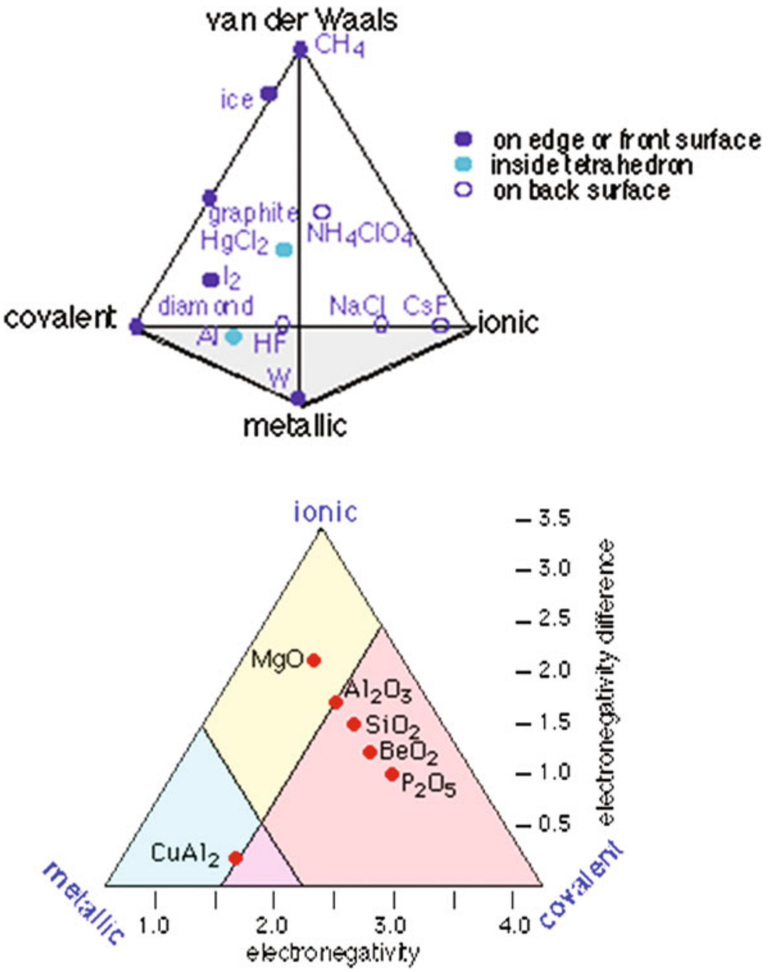


Fig. 5 Contribution of bonding types in common solids. Reproduced with permission from Jensen [2] © ACS 1998

In Fig. 5, we see a representation of the contribution of various kinds of bonding in common solids, represented in a tetrahedron. The corners represent 100 % contribution of a particular kind. Diamond is perfectly covalent, and tungsten metallic. The intermolecular interactions in gases (e.g., methane in the picture) is Van der Waals interaction. In graphite, the carbon atoms are arranged in strong sp2 covalent-bonded layers held together by weak Van der Waals interlayer forces. Thus in the materials bonding tetrahedral it is lying in-between the side having

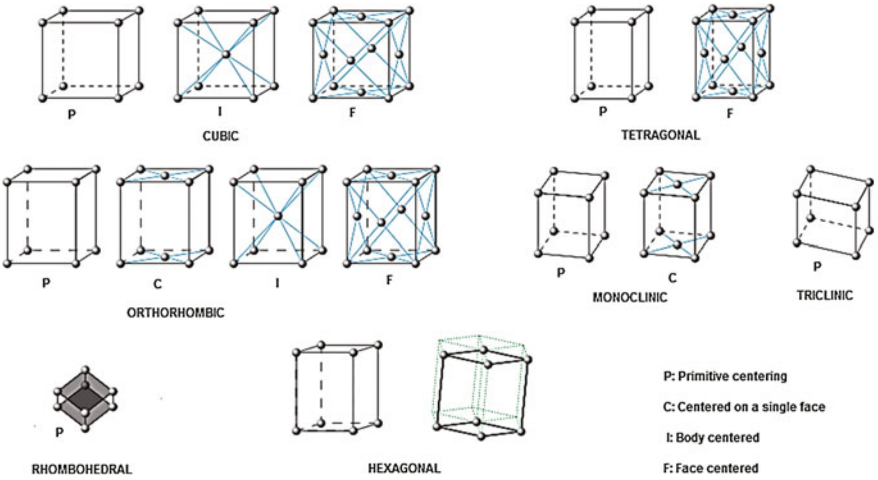


Fig. 6 The Bravais lattice. *P*, *C*, *I*, and *F* indicate primitive, base-centered (on a single face/base), body-centered, and face-centered crystals. The hexagonal unit cell is also separately highlighted

diamond and the methane (gas) at the two-end points. Ice is more toward the Van der Waals corner due to the dominance of H-bonding between the molecules in ice crystals. Not many solids are perfectly ionic; bonds between dissimilar atoms are somewhere in-between ionic and covalent and the percentage contribution due to the ionic bonding in such cases can be calculated as

$$I = 1 - \exp\left\{-\frac{(\chi_A - \chi_B)^2}{4}\right\} \tag{1}$$

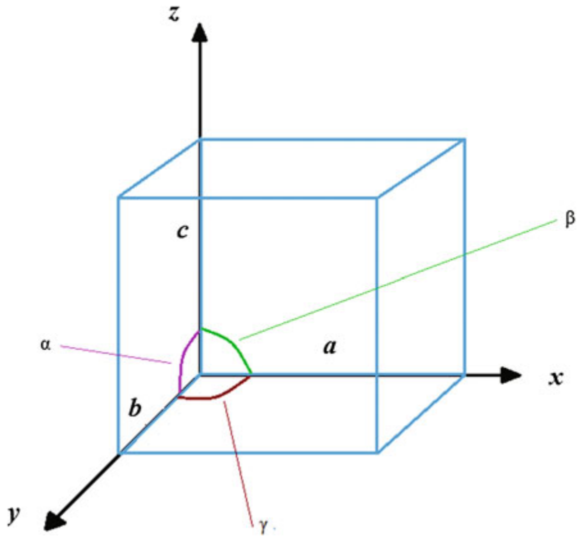
I represents the percentage of ionic character of the bond and $\chi_{A,B}$ represents the electronegativity of the constituent atoms A and B.

With such contributions from different kinds of bonding, the solids crystallize into various forms and configurations which can be geometry-wise classified into 14 types of Bravais lattices (Fig. 6).

There are 7 crystal systems, the cubic, tetragonal, orthorhombic, monoclinic, triclinic, rhombohedral, and the hexagonal. Further depending upon the position of atoms on the face or the in the center of the crystal, they can be distinguished as face-centered or body-centered systems.

The coordinate system of a 3D Cartesian lattice is described in terms of the sides of the unit cell *a*, *b,* and *c* and the angles between the planes α, β and γ (Fig. 7).

Fig. 7 The Cartesian system describing the crystal system



In terms of the coordinates, the seven systems can be easily distinguished as:

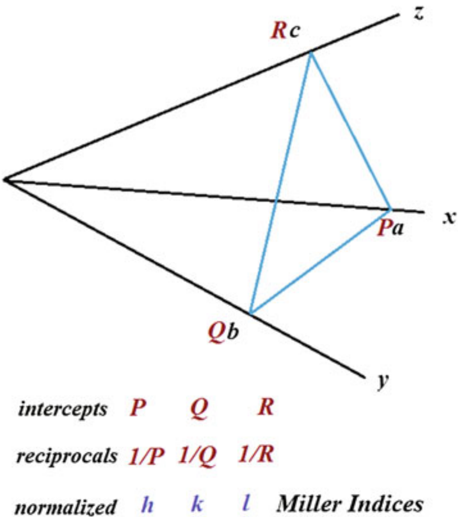
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$
Monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90° \neq \beta$
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90°$
Rhombohedral	$a = b = c$	$\alpha = \beta = \gamma \neq 90°$
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90°, \gamma = 120°$

Taking into account the variations depending upon body/face or base centered (one of the faces has an atom while others do not), there are 14 Bravais lattices to classify various crystallographic unit cells.

5 Miller Indices

A bulk solid contains a large number of crystals and hence a collection of set of parallel planes. An abstract way of representing such a family of planes with a particular orientation is the Miller Index. It is represented as (h k l) while the direction perpendicular to these set of planes is represented as [h k l] (Figs. 8 and 9).

Fig. 8 Schematic of Miller indices calculations



Say for example in an arbitrary lattice, the intercepts of a certain plane (representative of the set of planes parallel to it as well) are P_a Q_b and R_c. To calculate the miller indices, we take the intercepts and *P*, *Q,* and *R* and take their reciprocals and thereafter normalize them. For example, if the values are $P = 3$, $Q = 2$, and $R = 1$, then the miller indices would be (2 3 6).

If any axis has a negative intercept, the index is written with a bar sign above. If a plane is parallel to any axes, the intercept is considered at infinity and hence the reciprocal for that intercept is taken as 0.

In Fig. 9, we show for a cubic system various planes (shaded) and their Miller indices. The spacing between adjacent (parallel) planes with the same Miller indices ($h\ k\ l$) in an orthogonal lattice (angles $\alpha = \beta = \gamma = 90°$) is given by

$$d_{hkl} = \frac{1}{\sqrt{\left(\frac{h}{a}\right)^2 + \left(\frac{k}{b}\right)^2 + \left(\frac{l}{c}\right)^2}} \tag{2}$$

For rhombohedral and hexagonal lattices, it is useful to write the Bravais–Miller index ($h\ k\ i\ l$) where $i = -(h + k)$. Here *h*, *k,* and *l* are same as the Miller index, and *i* is a redundant index. Such a scheme in the hexagonal lattice makes the symmetries easier to understand.

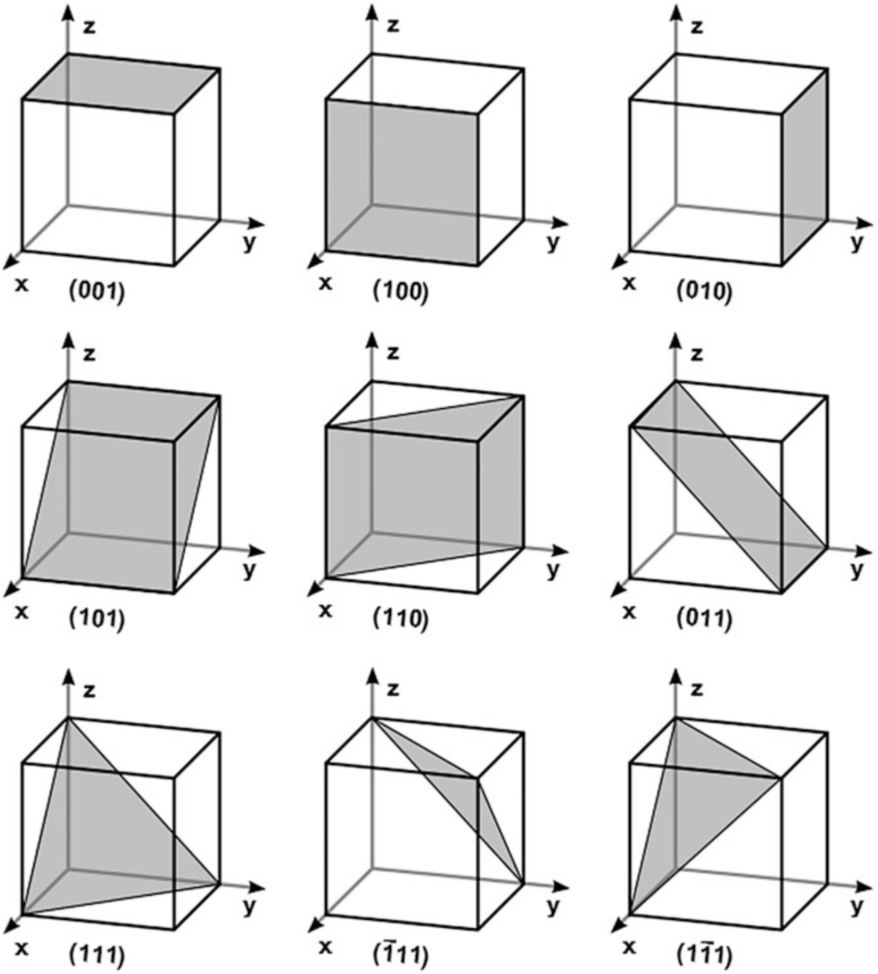


Fig. 9 Miller indices of various planes (shown as *shaded planes*) in a simple cubic system

6 The Reciprocal Lattice Space

A highly abstract concept in crystallography is the reciprocal lattice space. A 3D lattice can be considered a collection or a superset of large number of parallel set of planes. However, a set of planes can be represented by a single direction vector perpendicular to it. To identify the set of planes more uniquely, we can consider the length of this vector (perpendicular to the set of planes) to be reciprocal to that of the interplanar spacing. Thus the tip of the vector represents a point in space containing information about the orientation and the interplanar spacing among a certain set of planes. If we consider all sets of planes in the lattice, and represent them abstractly with such points, then we would get a collection of points having a certain correlation/arrangement among themselves. This collection of points also defines a lattice known as the reciprocal lattice of the crystal. The primitive reciprocal lattice is given in terms of the Bravais lattice $\vec{a}, \vec{b}$ and $\vec{c}$ as

$$\begin{aligned}\vec{A} &= 2\pi \frac{\vec{b} \times \vec{c}}{\vec{a} \cdot \left(\vec{b} \times \vec{c}\right)} \\ \vec{B} &= 2\pi \frac{\vec{c} \times \vec{a}}{\vec{a} \cdot \left(\vec{b} \times \vec{c}\right)} \\ \vec{C} &= 2\pi \frac{\vec{a} \times \vec{b}}{\vec{a} \cdot \left(\vec{b} \times \vec{c}\right)}\end{aligned} \tag{3}$$

The reciprocal lattice vector can also be expressed in terms of the Miller indices as

$$\vec{g}_{hkl} = h\vec{A} + k\vec{B} + l\vec{C}. \tag{4}$$

7 Bragg's Law

The crystal structure studies are usually carried out with X-ray diffraction or diffraction of monochromatic particle beams (e.g., mostly electron beam) (Fig. 10). Such radiations (or matter waves in case of particles) have wavelengths comparable to interplanar spacing in crystals and therefore can produce interesting diffraction patterns captured by film or electronic detectors, which after analysis can give detailed information about the crystal structure. This technique is commonly known

Fig. 10 Typical diffractometer schematic

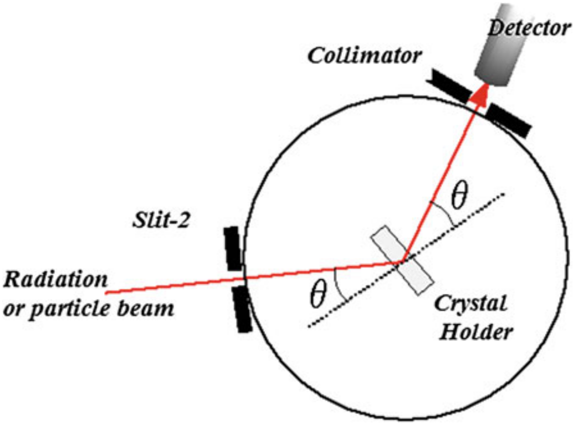
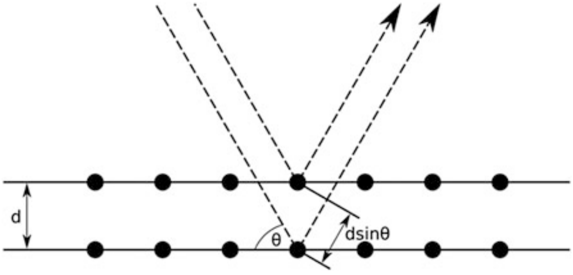


Fig. 11 Schematic of Bragg's Law



as crystallography. For such crystallographic studies, the atoms arranged in different planes of the crystal reflect X-rays/matter waves which interfere constructively or destructively. The condition for constructive interference is governed by the Bragg's law which states for an X-ray/matter wave of wavelength λ incident on a set of planes spaced by interplanar spacing d, the diffraction maxima (i.e., constructive interference of waves reflected by the crystals planes) occurs for the condition

$$2d \sin \theta = n\lambda \tag{5}$$

θ being the angle of incidence and n being the order of the diffraction maxima (also known as Bragg peaks) (Fig. 11).

8 Brillouin Zone

An important concept in crystallography is the Brillouin Zone, often shortened as BZ. It is a uniquely defined Wigner–Seitz primitive cell in reciprocal lattice space. The boundaries of this cell are given by planes related to points on reciprocal lattice (Fig. 12).

To find Brillouin zone of a certain lattice, we take its reciprocal lattice, draw vectors to the first nearest reciprocal lattice points, and then bisect them. The area defined by these perpendicular bisectors is known as the first BZ. If we further connect and similarly bisect the second nearest neighbors, we get an area defined by the bisectors which is the second BZ and so on for successive higher order zones. The zones may be different in shape to one another but always define the same area (2D case) or volume (3D case). From the crystallographic point of view, they are interesting as electron waves traveling through a solid are reflected back from the boundaries of the successive BZs.

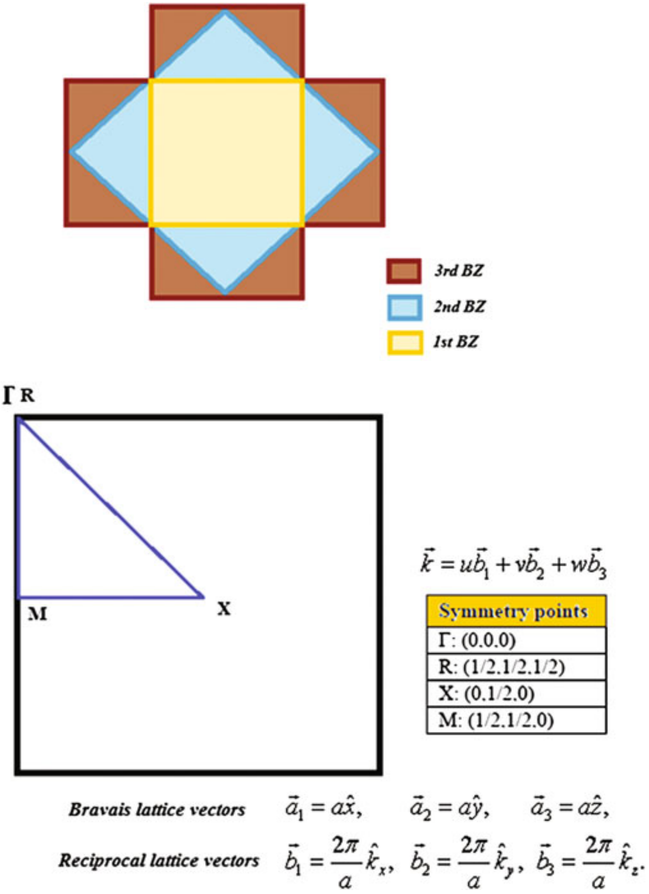

3rd BZ
2nd BZ
1st BZ
Γ R
M
X
$\vec{k} = u\vec{b}_1 + v\vec{b}_2 + w\vec{b}_3$
Symmetry points
Γ: (0.0.0)
R: (1/2.1/2.1/2)
X: (0.1/2.0)
M: (1/2.1/2.0)
Bravais lattice vectors $\vec{a}_1 = a\hat{x}, \quad \vec{a}_2 = a\hat{y}, \quad \vec{a}_3 = a\hat{z},$
Reciprocal lattice vectors $\vec{b}_1 = \frac{2\pi}{a}\hat{k}_x, \ \vec{b}_2 = \frac{2\pi}{a}\hat{k}_y, \ \vec{b}_3 = \frac{2\pi}{a}\hat{k}_z.$

Fig. 12 Brillouin Zones of the 2D square lattice. High symmetry points in the Brillouin Zone of simple cubic crystal (projection view)

9 Atomic Packing Factor

Not all crystal systems have the same order of atomic layer stacking structures. Some structures are more densely packed and some are loosely packed. The atomic packing factor is a measure of how densely a structure is packed with atoms and is defined as the ratio of the volume of the atoms in a unit cell to the total volume of the unit cell (Fig. 13).

$$\text{Atomic packing factor (APF)} = (\text{Volume of atoms in unit cell})/(\text{Volume of the unit cell})$$

In SCC unit cell, there are eight atoms in the corners. Each corner atom is shared among eight unit cells.

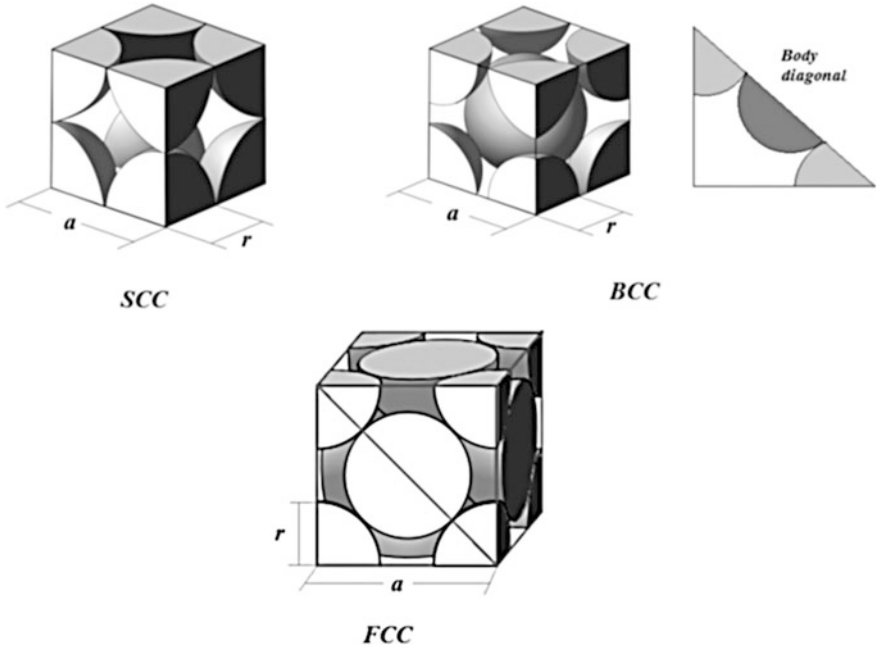


Fig. 13 Atomic packing in cubic crystals

Thus, number of atoms in a unit cell is $8 \times \frac{1}{8} = 1$.

For the above figure, along the side of the cube $a = 2r$.

Thus the atomic packing factor in this case is

$$\begin{aligned} \text{APF} &= \frac{\text{No of Atomsin unit cell} \times \text{Vol of atom}}{\text{Vol of unit cell}} \\ &= \frac{\left(1 \times \frac{4}{3}\pi r^3\right)}{a^3} = 0.52 \end{aligned}$$

In BCC unit cell, there are eight atoms in the corners and one atom in the center of the unit cell. Each corner atom is shared among eight unit cells. The central atom is possessed by only one unit cell. Thus, the number of atoms in a unit cell is $1 + 8 \times \frac{1}{8} = 2$.

From figure, the number of atoms along the diagonal of the body is $a = \frac{4}{\sqrt{3}} r$.

Thus the atomic packing factor in this case is

$$\text{APF} = \frac{\left(2 \times \frac{4}{3}\pi r^3\right)}{a^3} = 0.68$$

In FCC unit cell, there are eight atoms in the corners and six atoms on faces of the unit cell. Each corner atom is shared among eight unit cells and each face atom is shared between two unit cells. Thus, the number of atoms in a unit cell is $8 \times \frac{1}{8} + 6 \times \frac{1}{2} = 4$.

In FCC along the face diagonal $a = 2\sqrt{2}r$.

Thus the atomic packing factor in this case is

$$\mathrm{APF} = \frac{\left(4 \times \frac{4}{3}\pi r^3\right)}{a^3} = 0.74.$$

10 Some Important Crystal Structures

10.1 Rocksalt (NaCl) Structure

The rocksalt or NaCl structure is a very common structure among binary ionic compounds. NaCl is a structure consisting of two interpenetrating FCC lattices, where larger chlorine atoms occupying the face-centered cubic position with the smaller sodium ions are placed in the octahedral void (outline indicated in blue) between them. The sodium ions also form an FCC lattice on their own (Fig. 14).

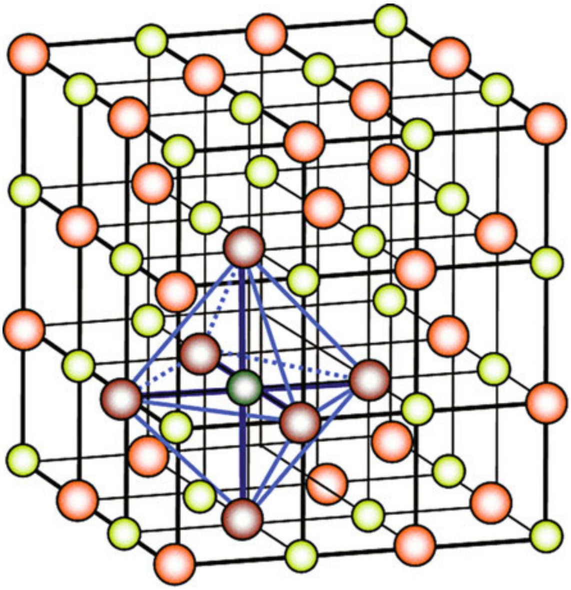


Fig. 14 Rocksalt structure

10.2 Diamond Cubic Structure

Diamond is one of the allotropes of carbon and is the hardest material known to man. It is a covalent-bonded solid arranged in an FCC space lattice with two atoms per lattice point. It is a covalent sp3-bonded structure with a coordination number (min equidistant neighbors) of four.

In Fig. 15, the projected structure is also shown for clarity. The fractional numbers corresponding to the atoms indicate the height of that atom considered from the base of the cell, where the height of the entire cell is considered as unity. The atoms not marked with any numbers (at the four corners and the center) indicate the overlapping base and the top layers of atoms at fractional coordinates 0 and 1. This structure has a low packing density of about 0.34.

Silicon and Germanium are other two very important electronic materials that crystallize in the diamond cubic structure.

10.3 Zinc Blende

The Zinc Blende structure is analogous to the diamond cubic structure but consisting of dissimilar atoms and having a slightly higher packing fraction than diamond (depending upon the material). Many compound semiconductors GaAs, InSb, SiC, CdTe etc. crystallize in this form. In ZnS structure, the two atom types constitute two interpenetrating FCC lattices, but their relative positioning is different from the rocksalt structure (Fig. 16).

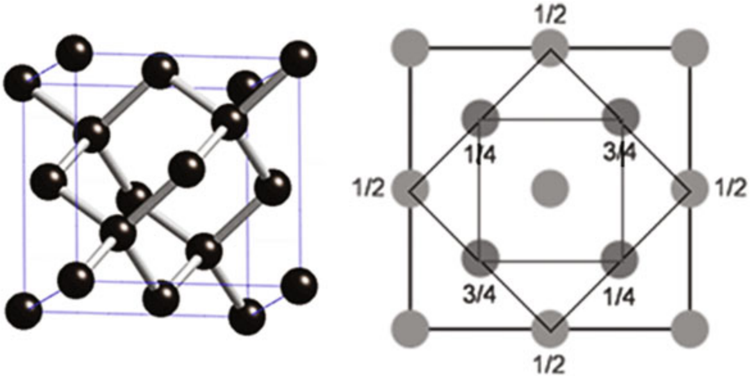


Fig. 15 Diamond cubic unit cell and the in-plane projection of the structure (*numbers* indicate height from base of the cell)

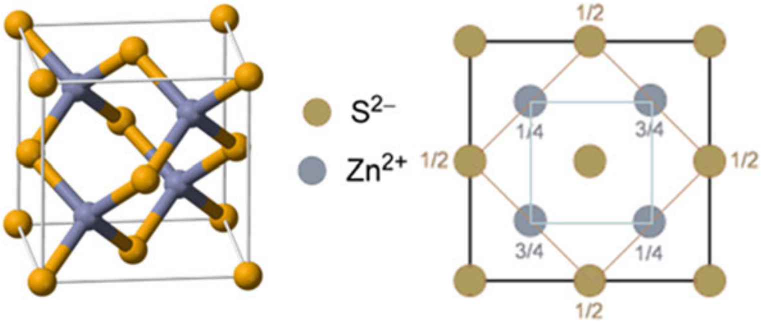
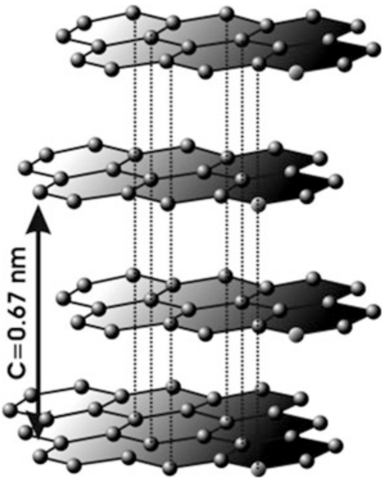


Fig. 16 ZnS unit cell and the in-plane projection of the structure (*numbers* indicate height from base of the cell)

10.4 Graphite

Graphite is another allotrope of carbon having a layered structure of sp2 bonded sheets held together by weak Van der Waals bonds (Fig. 17). The individual sheets are of honeycomb structure with C–C sp2 bond length 0.142 nm and all the atoms lying in one plane. The interlayer separation is about 0.335 nm and the unit cell c axis has length 0.67 nm [1–7]. Due to the weak bonds holding these layers together, the layers are able to slide over one another upon application of shearing stress and thus graphite is able to provide good lubricating action in certain applications. Depending on the relative stacking of the layers, graphite can be alpha (hexagonal) or beta (rhombohedral) phase.

Fig. 17 The structure of graphene



10.5 Graphene

The individual freestanding honeycomb layers of graphite when chemically/mechanically exfoliated from the bulk show remarkably different properties than the bulk graphite. First, in such a material the carriers are effectively confined in one direction (e.g., z) and are free to move in the other two (x–y). This gives rise to quantum confinement in the (say z) direction. The carriers in such a material have only 2 degrees of freedom and therefore such materials are known as 2D materials. Graphene, Silicene, Germanene, monolayer transition metal dichalcogenides (TMDs), and phosphorene etc., fall in this category of materials [1, 4, 7–47].

The first free-standing graphene was discovered in 2004 by A.K. Geim and K. Novoselov, [4, 13] Fig. 18 by the famous scotch tape method (also known as mechanical exfoliation) from bulk graphite. Since then, several other methods including chemical exfoliation, chemical vapor deposition, micromechanical cleavage, epitaxy and reduction of graphene oxide have been also adopted for graphene fabrication. Owing to the discovery of a new class of materials, Geim and Novoselov were awarded the 2010 Nobel Prize in Physics [13].

Graphene, a 2D material is an sp2-bonded honeycomb of carbon atoms lying in the same plane. The C–C bond length in graphene is 0.142 nm and the unit cell of graphene is made up of two carbon atoms (labeled A and B) in the Fig. 18. The unit cell is an equilateral parallelogram with sides 0.246 nm. The primitive vectors are given by

$$\begin{aligned} \vec{u}_1 &= a\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right) \\ \vec{u}_2 &= a\left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right) \end{aligned} \tag{6}$$

where a is the C–C bond length. The reciprocal lattice of graphene is also honeycomb in nature, with primitive vectors

$$\begin{aligned} \vec{v}_1 &= \frac{2\pi}{a}\left(\frac{1}{\sqrt{3}}, 1\right) \\ \vec{v}_2 &= \frac{2\pi}{a}\left(\frac{1}{\sqrt{3}}, -1\right) \end{aligned} \tag{7}$$

The high symmetry points of the hexagonal first Brillouin Zone of graphene are $\Gamma(0,0,0), M\left(0,\frac{1}{2},0\right)$ and $K\left(\frac{1}{3},\frac{1}{3},0\right)$ points. Primed points are equivalent high symmetry points (Fig. 19).

The most interesting property lies in the band structure of graphene. The valence and conduction bands of graphene touch each other at the six K points of the Brillouin zone making it a zero bandgap material and giving it exceptional electrical and thermal conduction properties. The dispersion ($E-k$) relation at those points of the BZ are linear in nature [4, 13]

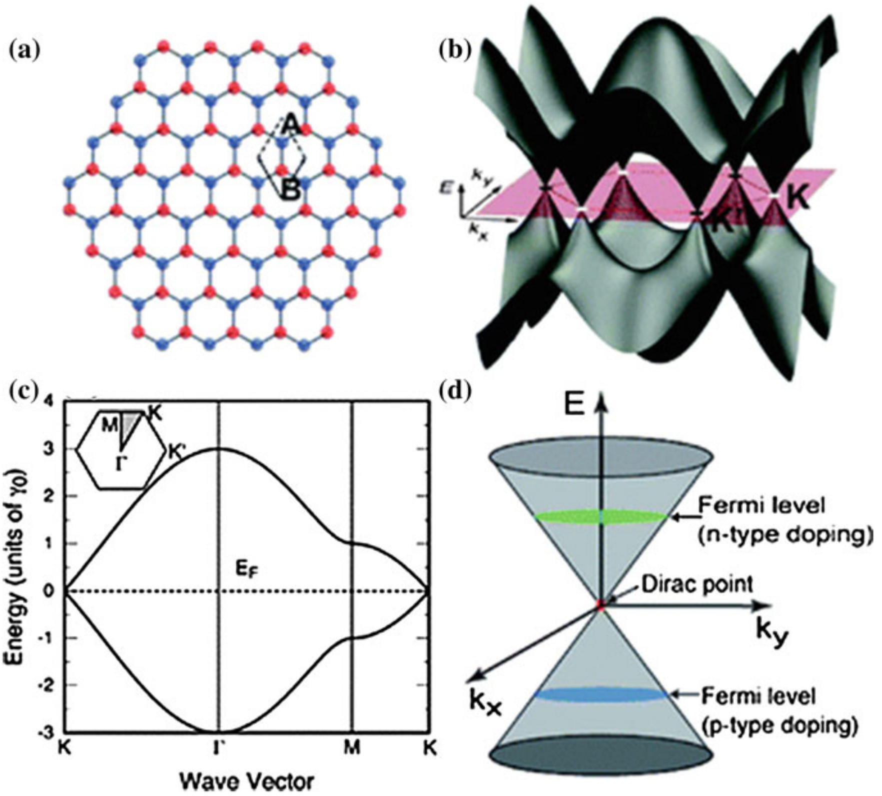
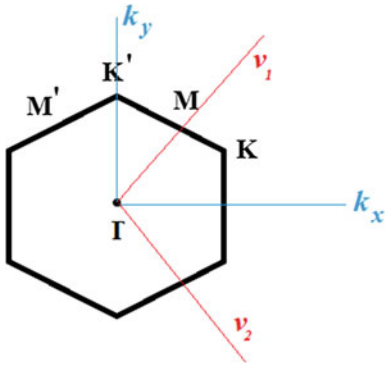


Fig. 18 **a** Hexagonal honeycomb lattice of graphene with two atoms (**a** and **b**) per unit cell. **b** The 3D band structure of graphene. **c** Dispersion of the states of graphene. **d** Approximation of the low-energy band structure as two cones touching at the Dirac point. The position of the Fermi level determines the nature of the doping and the transport carrier. Reprinted with permission from Avouris et al. [3]. © ACS 2010

Fig. 19 Brillouin Zone and high symmetry points of graphene



$$E(k) = \pm\hbar\upsilon_F\sqrt{k_x^2 + k_y^2} \tag{8}$$

$\hbar$ is the reduced Planck's constant and υ_F is the Fermi velocity k being the wave vector. The fact that the valence (VB) and conduction bands (CB) touch each other at these K points of the BZ (which are also known as Dirac points for graphene) lead to the advantage of charge pumping from VB to CB without having to overcome any bandgap. Therefore, intrinsic graphene has a very high amount of carrier density given by [4, 13]

$$n_i = \frac{\pi}{6}\left(\frac{\kappa_B T}{\hbar\upsilon_F}\right)^2 \tag{9}$$

which corresponds to a value of $\sim 10^{10}/\mathrm{cm}^2$ at room temperature. Such huge availability of carriers makes graphene an excellent conductor. Free-standing sheets of graphene show doping-independent high carrier mobility of over 200 $\mathrm{cm}^2\ \mathrm{V}^{-1}\ \mathrm{s}^{-1}$ (carriers in graphene are considered to be massless Dirac Fermions). Further, it is optically transparent and highly flexible but stronger than steel (Fig. 20).

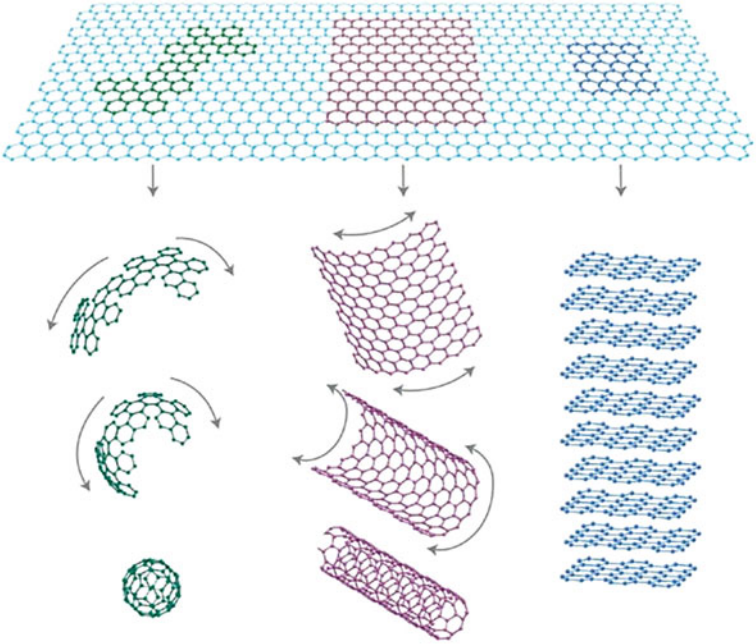


Fig. 20 Graphene is a 2D building material for carbon materials of all other dimensionalities. It can be wrapped up into 0D buckyballs, rolled into 1D nanotubes or stacked into 3D graphite. Reprinted with permission from Geim and Novoselov [4], © Nature Publishing Group 2007

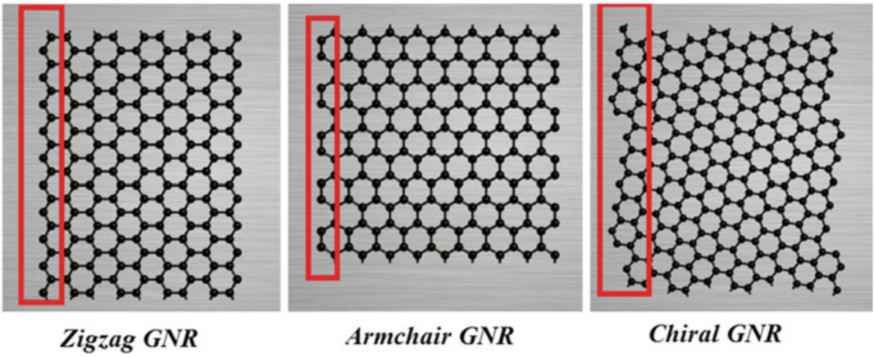


Fig. 21 Zigzag, Armchair, and Chiral Graphene Nanoribbon (GNR)

11 Graphene Nanoribbon

If graphene sheets are cut into thin strips (mostly by chemical patterning), then in addition to the confinement in the vertical *z* direction, a lateral confinement in *x* or *y* is also introduced. The degree of this confinement depends upon the width of the thus fabricated nanoribbon and also the nature of the cut of the edges (Figs. 20 and 21).

Depending on the cut on the sides of the ribbon, they are categorized as zigzag, armchair and chiral (mixture of armchair and zigzag edges) nanoribbons (Fig. 21).

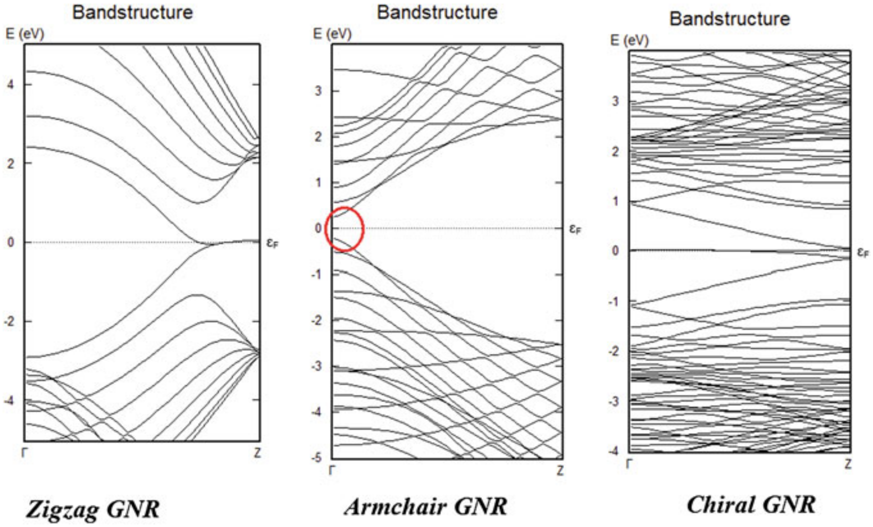


Fig. 22 Band structures for different kinds of GNR

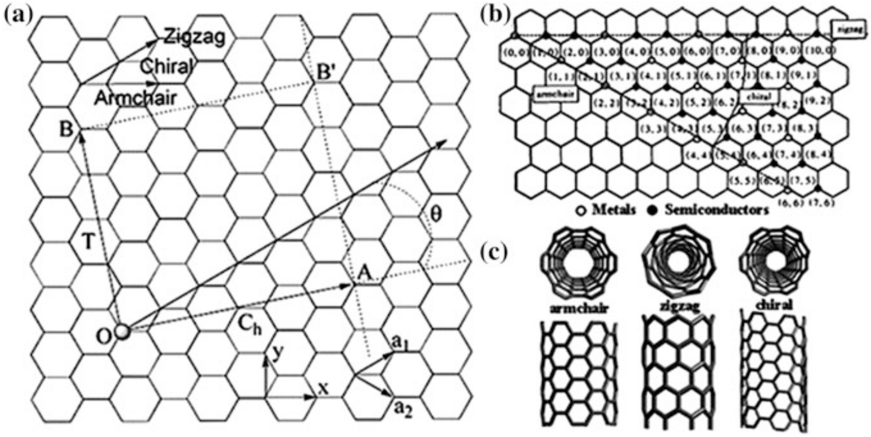


Fig. 23 **a** Schematic plot of the chiral vector (C_h) in a graphene 2D lattice. **b** The relationship between integers (n, m) and the metallic or semiconducting nature of nanotubes. **c** The structure of "armchair", "zigzag", and "chiral" nanotubes. Reprinted with permission from Zhang et al. [5]. © Royal Society of Chemistry 2011

For the electronic properties, the Zigzag GNR is always metallic while the armchair GNR may be metallic or semiconducting (depending on width) and the Chiral GNR band structure depends on its chirality (Fig. 22). The chiral vector is defined as [1]

$$C_h = n\vec{a}_1 + m\vec{a}_2 \tag{10}$$

where $\vec{a}_1$ and $\vec{a}_2$ are the basic vectors of the graphene unit cell. n, m are positive integers $(0 \leq m \leq n)$. If the ribbon edge is cut along a chiral vector $(n, 0)$ it becomes a zigzag nanoribbon and if the cut is along (n, n) it results in an armchair nanoribbon. If $n \neq m$ then it is a (n, m) chiral nanoribbon with electronic properties in-between the pure armchair and the pure zigzag GNRs (Fig. 23).

12 CNT

Carbon nanotubes or CNTs are a class of 1D material formed by sheets of graphene rolled into hollow tubes of small (0.5–5 nm) diameter. They were first discovered by Sumio Ijima in 1991. These were first observed in samples of soot. Since then, efficient synthesis methods such as laser vaporization synthesis and arc-discharge methods have been developed for fabricating CNT. CNT can be classified as single-walled CNT (SWCNT) or multiwalled CNT (MWCNT). Multiwalled CNT may be made up of a number of coaxial CNTs of varying diameters (also known as the Russian doll type), or a single sheet rolled spirally into a tube of large diameter (this kind is also known as the more technically appropriate term carbon nano-scroll) (Fig. 24).

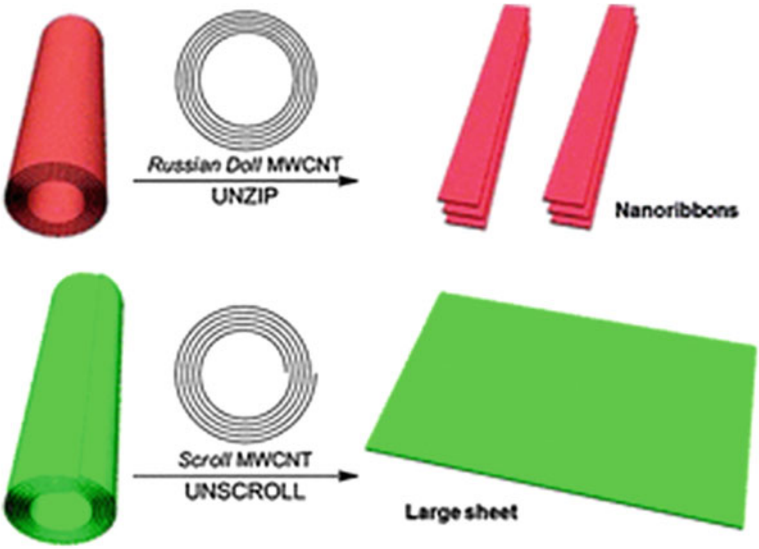


Fig. 24 Russian Doll and scroll-type MWCNTs. Reprinted with permission from [6] © Royal Society of Chemistry 2013

Depending on the direction of rolling, CNT (here as we refer to CNT we mean individual SWCNTs) cab be classified as armchair, zigzag, or chiral. If we roll the sheet with its axis along the translational vector $\vec{T}$, which is perpendicular to the chiral vector C_h, then the nanotube can be called as armchair CNT because $n = m$ if $m = 0$ its an zigzag CNT, and if $n \neq m\, n, m > 0$ then it is a (n, m) chiral CNT. The diameter of an individual SWCNT is given by [1]

$$d_{CNT} = a\sqrt{n^2 + nm + m^2} \tag{11}$$

Another important parameter known as the chiral angle (which is the angle between the Chiral $\vec{C}_h$ vector and the primitive vector $\vec{a}_1$) is given by [1]

$$\theta = \frac{2n + m}{2\sqrt{n^2 + nm + m^2}} \tag{12}$$

Owing to the hexagonal structure of graphene $0^\circ \leq \theta \leq 30^\circ$. If $\theta = 0^\circ$ then the CNT is zigzag, if $\theta = 30^\circ$ it is armchair, and all values in between are chiral CNTs.

As for the electronic properties, all armchair nanotubes are metallic (valence and conduction bands cut each other), while zigzag and chiral CNTs may be metallic/semiconducting based on the n, m values (Fig. 25). For the zigzag CNTs if $n - 1$ is an integral multiple of 3 then it is a semiconducting one with bandgap given by [1]

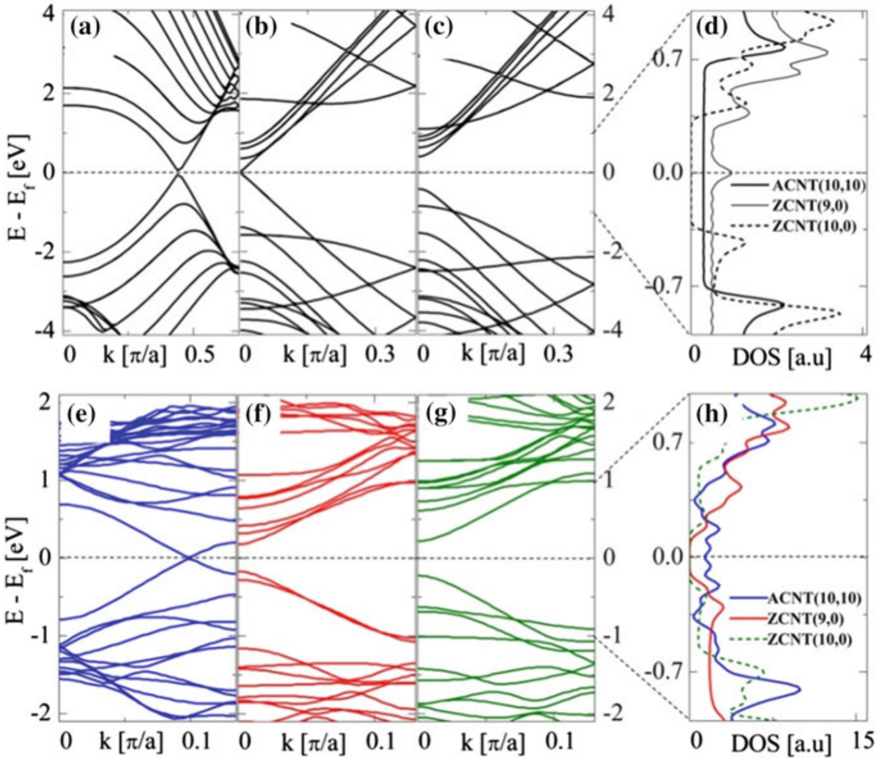


Fig. 25 Band structures in pristine CNTs: **a** ACNT (10,10), **b** ZCNT (9,0), **c** ZCNT (10,10); **d** DOS. Reprinted with permission from Baskin et al. [6] © Nature Publishing Group 2011

$$E_g \approx 2\gamma \frac{a_{C-C}}{d_{CNT}} \tag{13}$$

In the above equation, $\gamma = 3.1$ eV and $a_{C-C} = 0.142$ nm. For all other zigzag CNTs, the nature is metallic. The bandgap of the chiral CNTs are usually obtained from Kataura plots.

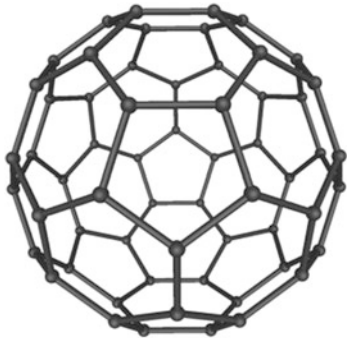


Fig. 26 A C60 molecule. (Pic courtesy WikiMedia Commons)

13 Fullerene

Fullerenes or Buckminsterfullerene, which are commonly known as buckyballs are one of the first carbon-based nanomaterials discovered in 1985 (Fig. 26). It is a football resembling structure comprising of hexagons and pentagons with the carbon atoms lying at the corners. The chemical formula for fullerene is C_{60} (60 represents the number of C atoms). Fullerene-like materials with smaller number of atoms 20, 24, 26 etc. are also possible; however, the structure in such cases may not be the same and there are shape anisotropies. The C_{60} structure is formally known as truncated icosahedral (also known as the Geodesic dome). In a fullerene, the electrons are confined in all three directions and hence it represents a 0D material. The fullerene energy states are similar to that of a quantum dot and consist of discrete energy levels.

14 Other 2D Materials

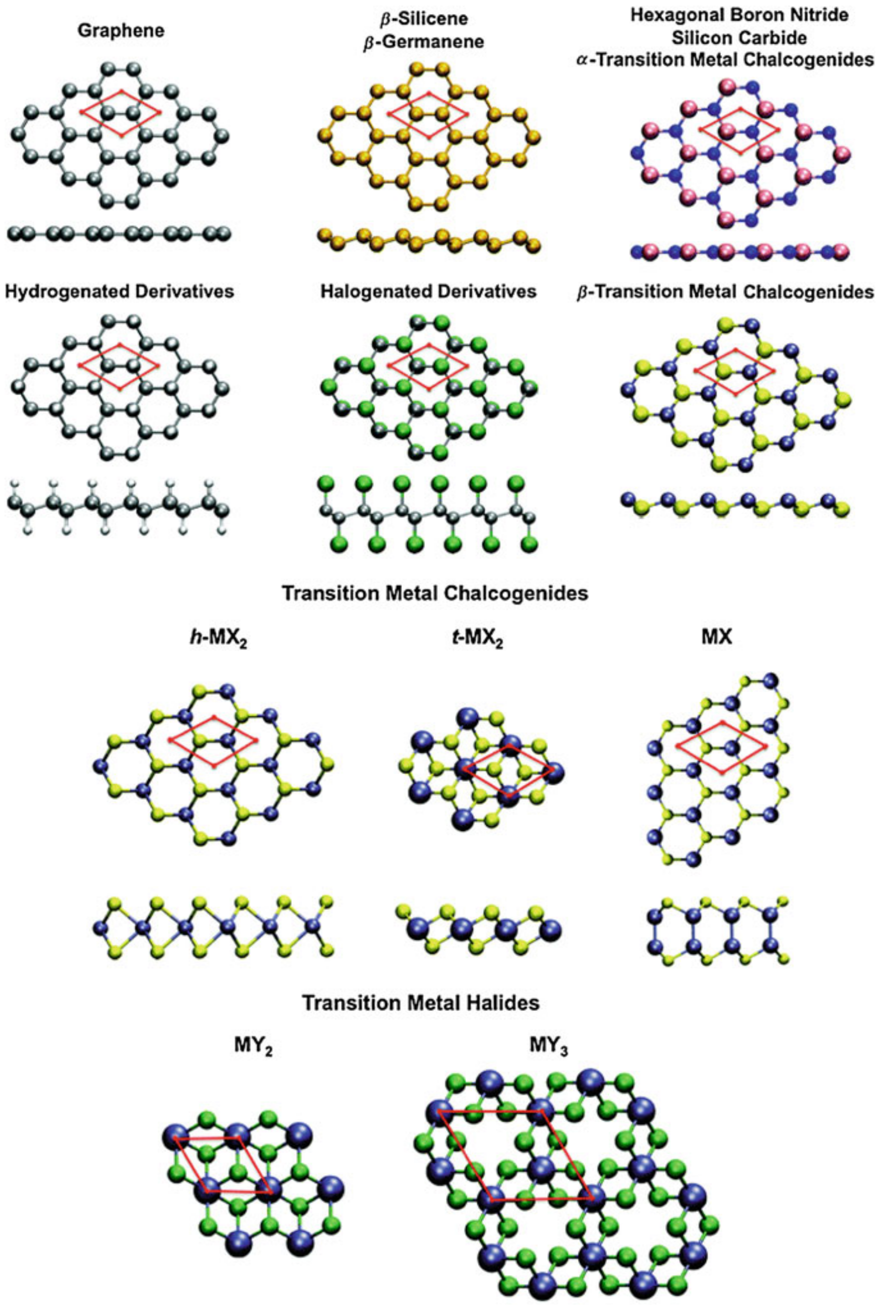
With the Si MOS technology heading toward the scaling limits, efforts are on worldwide to search for an alternative technology to serve as its replacement in the near future. Increasing the performance and the energy efficiency of ICs by scaling is facing severe roadblocks even with strained Si channels and multigate Si MOSFET structures. In order to achieve the “deeper nanoscale” regime of 10 nm and less, alternatives to Si CMOS must be explored. Of the various classes of channel materials under research, the 2D materials like graphene are of great significance [9–17]. The chief advantages of 2D channels over bulk channels in MOSFET are the suppression of the drain-induced barrier lowering (DIBL) which is a major problem in nanoscale MOSFETs and the better electrostatic control over the channel (resulting in low power consumption). Also for 2D materials like graphene, the extremely high carrier mobilities and very high saturation velocities can provide for good analog applications like amplifiers. Also high cut-off frequencies of graphene and other 2D materials could lead to excellent RF applications. The 2D materials possess optical transparency and mechanical flexibility which are also suited for optoelectronics applications and flexible ICs. A big advantage of 2D nanosheets as channel material over other low-dimensional materials (like nanotubes, nanowires etc.) lie in the ease of fabricating MOS structures with them. Such ultra-thin layers are very much compatible to standard CMOS fabrication techniques. Thus 2D channel material-based next-generation MOSFET are an emerging and a very important area of research, which could shape the future of the VLSI industry and consumer electronics in the years to come [8–47].

Since the discovery of graphene in 2004, there has been a huge interest generated in the so-called 2D materials. Due to the remarkable electrical properties of graphene such as doping-independent mobility, superior transport properties by Dirac Fermions, the optical transparency, and compatibility with flexible electronics, it has become a highly researched material for electronic applications and a potential candidate for future CMOS technology. At the same time, a number of other 2D materials have come to the fore as major candidates as alternative channel materials. Materials like Silicene (single layer Si), Germanene (monolayer Ge), Transition Metal Dichalcogenide (e.g., MoS_2), topological insulators have shown promising electrical properties suitable for device and interconnect applications [8–47].

15 2D MoS_2 and Other Transition Metal Dichalcogenides

MoS_2 is a well-known transition metal dichalcogenide (a class of layered materials having general formula MX_2: M = Mo, W; X = S, Se, Te) built of van der Waals bonded S-Mo-S units. The layers are composed of an atomic plane of Mo sandwiched between two atomic planes of S in a trigonal prismatic arrangement. The unit cell of MoS_2 is hexagonal having six atoms, the lattice parameters are $a = 3.160$ Å, $c = 12.294$ Å, and d (Mo–S) = 4.21 Å and the interlayer spacing $\Delta = 2.975$ Å. These sheets of loosely bound MoS_2 are quite similar to the honeycomb monolayers of graphene in graphite, and are responsible for the lubricating properties of MoS_2. The bulk MoS_2 is an indirect bandgap semiconductor with an intrinsic bandgap of 1.2 eV. However, the monolayers (S–Mo–S sheets) display rather different electrical properties like direct bandgap of 1.8 eV, optical transparency, enhanced photoluminescence etc. when isolated from bulk MoS_2. Being a layered material, atomically thin layers of MoS_2 can be obtained by standard processes such as micromechanical cleavage (scotch tape mechanical exfoliation) and liquid exfoliation. The process of obtaining ultrathin layers from bulk MoS_2 by mechanical peeling off had been reported way back in 1965 by Frindt. While in 1986 Joensen, Frindt, and Roy Morrison reported a novel process of intercalation of 2H-MoS_2 powder with lithium, followed by an interaction with water and ultrasonication, yielding monolayer MoS_2 [24–47]. Figs. 27, 28, 29 and 30 show the lattice and band structures of various 2D materials.

Till the late 2000s, the focus of studies on such monolayer MoS_2 was related to formation of inclusion compounds, catalytic effects, and tribological studies. It was only after the regeneration of applied interest in the field of 2D materials in 2004, that a number of interesting studies on the electronic and optical properties of the monolayer MoS_2 came to the fore. In the recent years, quite a few experimental studies on electrical and optical properties of MoS_2 monolayers have been carried out by various researchers around the world. In their study on the low temperature photoluminescence properties of 2D MoS_2, Splendianiet et al. and Kornet et al. have observed interesting PL properties and the appearance of low-energy PL peaks, at low temperatures which point to the presence of bound exciton states in such a material. Such properties could lead to the development of fast photoconductive switches based on MoS_2. As for the electrical properties it has been found by the two probe and the four probe measurements by Ayari et al. that 30-nm thick MoS_2 suspended on SiO_2 showed a rather limited mobility in the range of


Graphene
β-Silicene
β-Germanene
Hexagonal Boron Nitride
Silicon Carbide
α-Transition Metal Chalcogenides
Hydrogenated Derivatives
Halogenated Derivatives
β-Transition Metal Chalcogenides
Transition Metal Chalcogenides
h-MX2
t-MX2
MX
Transition Metal Halides
MY2
MY3

◀ **Fig. 27** Structure of atomically thin 2D materials: graphene, silicene, germanene (and their derivatives), silicon carbide (SiC), hexagonal boron nitride (*h*-BN), and transition metal chalcogenides (TMC) (*top*). Ultrathin 2D materials such as transition metal chalcogenides (TMC) (*center*) and halides (TMH) (*bottom*). Unit cells are depicted in *red*. Color code: metals in *ice blue*, halides in *green*, chalcogenides in *yellow*, nitrogen in *blue*, carbon in *gray*, silicon and germanium in gold, boron in *pink*, and hydrogen in *white*. Reprinted with permission from Miro et al. [7]. © Royal Society of Chemistry 2014 (Color figure online)

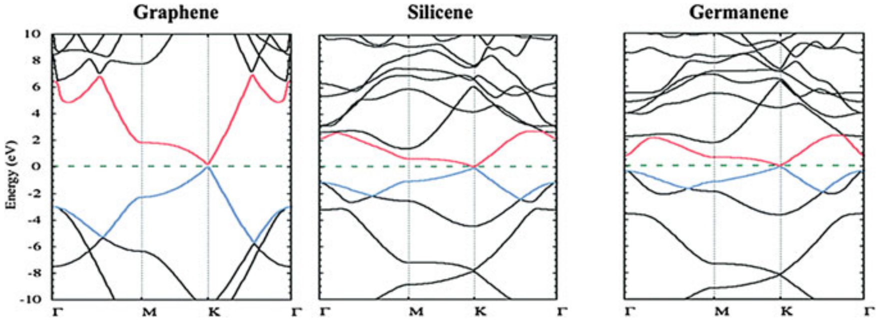


Fig. 28 Band structures of graphene (*left*), silicene (*center*), and germanene (*right*). Valence and conduction bands are highlighted in *blue* and *red*, respectively. The Fermi levels (*horizontal green dashed lines*) have been shifted to 0 eV. All band structures at PBE-D3 level of theory. Reprinted with permission from Miro et al. [7]. © Royal Society of Chemistry 2014

10–50 cm^2 V^{-1} s^{-1}. This was better than the earlier shown mobilities in the range of 0.5–3 cm^2 V^{-1} s^{-1}. However, it was still a roadblock. To overcome this issue, Radisavljevic et al. have employed dielectric screening in the presence of a high-*k* material like HfO_2 to successfully boost the mobility of 2D MoS_2 to 200 cm^2 V^{-1} s^{-1} and more (which is as good as Graphene). The device fabricated at EPFL, Switzerland (by the group of Prof. AndrasKis) has shown considerable advantage over the conventional Si MOS and the graphene FET. Such a MoS_2 MOSFET has 10^5 times less energy consumption in standby mode compared to the conventional Si MOSFET. Also MoS_2 monolayer possesses a bandgap of 1.8 eV which is highly suited for switching applications. In another work, H. Liu and Prof. P.D. Ye at the Nanotechnology Center, Purdue University, USA, have fabricated dual-gate MOSFET with 2D MoS_2 channel with ALD-grown Al_2O_3 top-gate, displaying very high ON/OFF ratios of 10^8 and extremely high electron mobility of 500 cm^2 V^{-1} s^{-1}. Such mobility is about 2.5 times higher than that of graphene FETs.

Apart from 2D MoS_2, multilayer flakes of the molybdenite are also being investigated for MOSFET applications. The 15-nm thick MoS_2 flake-based electric double layer transistors (EDLT) have been recently fabricated by the group of Prof. Yoshihiro Iwasa at University of Tokyo, Japan. They have reported the detection of ambipolar transport in such EDL FETs, which makes the material promising for new device functionalities like light emitting devices. The thin, transparent MoS_2 channel is also highly suited for optoelectronic applications and flexible electronics.

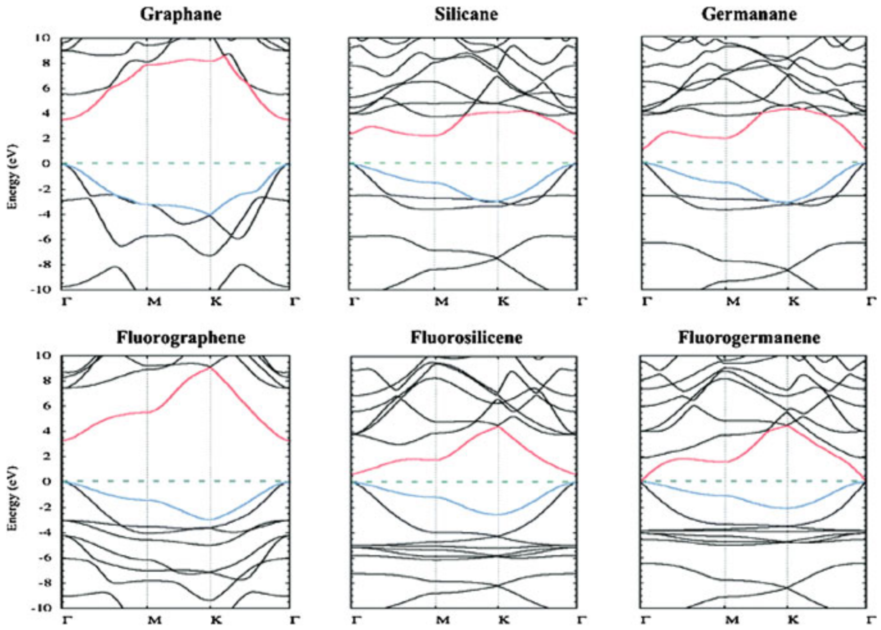


Fig. 29 Band structures of hydrogenated (*top*) and fluorinated (*bottom*) derivatives of graphene (*left*), silicene (*center*), and germanene (*right*). Valence and conduction bands are highlighted in *blue* and *red*, respectively. The Fermi levels (*horizontal green dashed lines*) have been shifted to 0 eV. All band structures at PBE-D3 level of theory. Reprinted with permission from Miro et al. [7]. © Royal Society of Chemistry 2014

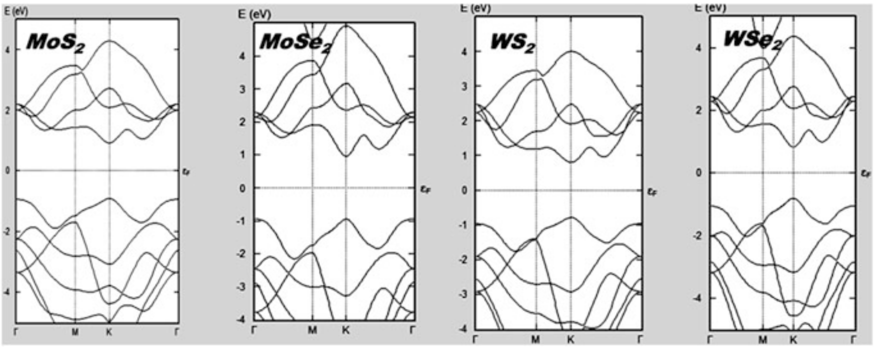


Fig. 30 Band structures of some 2D TMD materials

Therefore, it is of considerable interest to focus on the application of monolayer as well as multilayer MoS_2 as an alternative channel material for CMOS application. Apart from the 2D monolayer, MoS_2 nanotubes and nanoribbons could also offer novel transport properties suited for other electronic devices (e.g., Tunnel FET).

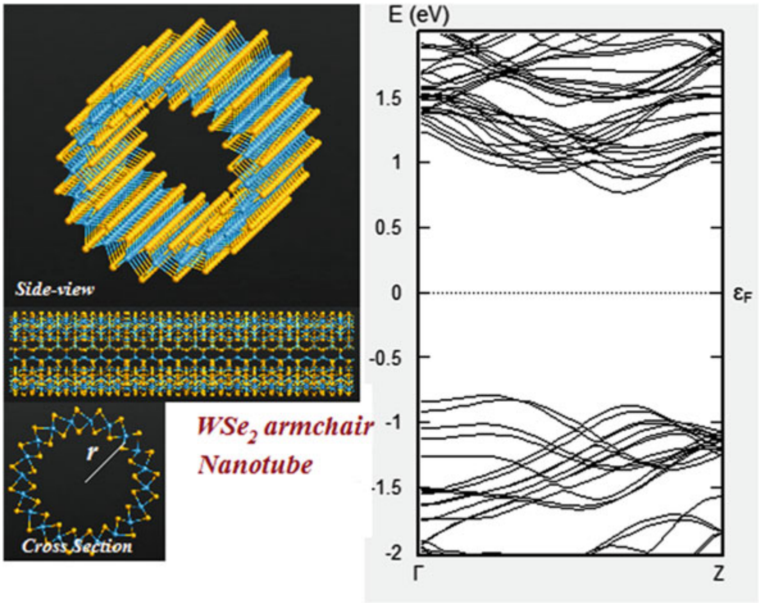


Fig. 31 Structure and simulated tight-binding band structure of WSe_2 armchair nanotube of ~2.5 nm diameter showing a semiconducting nature

The transition metal dichalcogenides (TMD or MX_2) all have the same crystal structure, built up of hexagonal X-M-X monolayers having intrinsic bandgaps of 1.1–2.0 eV depending upon the material) and when isolated by chemical or mechanical exfoliation, they all demonstrate nonzero bandgap in their 2D-sheet form. In recent studies, other (MX_2), in particular tungsten-based TMDs (like WSe_2, WSe_2), have shown superior performance to molybdenum-based TMDs (like MoS_2, $MoSe_2$) in terms of electronic mobility and MOS device performance. Nanotubes of WS_2 have also been realized for FET applications recently (Fig. 31). These developments in TMDs have created a great opportunity for future electronic devices [4–8], [24–47].

References

1. R. Saito, G. Dresselhaus, M.S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (World Scientific, Singapore, 1998)
2. W.B. Jensen, Logic, history and the chemistry textbook. J. Chem. Educ. **75**, 817–828 (1998)
3. P. Avouris et al., Nano Lett. **10**(11), 4285–4294 (2010)
4. A.K. Geim, K.S. Novoselov, Nat. Mater. **6**, 183–191 (2007)
5. H. Zhang et al., Chem. Soc. Rev. **40**, 1324–1336 (2011)
6. C.H. An-Wong et al., Phys. Chem. Chem. Phys. **15**, 7755–7759 (2013)
7. P. Miro et al., Chem. Soc. Rev. **43**, 6537–6554 (2014)

8. A. Baskin et al., Scientific Reports 1, **36** (2011). doi:10.1038/srep00036
9. International Technology Roadmap for Semiconductors (ITRS) Report 2011: Emerging Research Devices
10. International Technology Roadmap for Semiconductors (ITRS) Report 2011: Emerging Research Materials
11. K.S. Novoselov et al., Electric field effect in atomically thin carbon films. Science **306**(5696), 666–669 (2004)
12. K.S. Novoselov et al., Two-dimensional atomic crystals. Proc. Natl. Acad. Sci. USA **102**(30), 10451–10453 (2005)
13. A.K. Geim, K.S. Novoselov, The rise of graphene. Nat. Mater. **6**(3), 183–191 (2007)
14. S. Cahangirov, M. Topsakal, E. Aktürk, H. Şahin, S. Ciraci, Two- and One-dimensional honeycomb structures of silicon and germanium. Phys. Rev. Lett. **102**(23), 1–4 (2009)
15. E. Scalise, M. Houssa, G. Pourtois, V. Afanas'ev, A. Stesmans, Strain-induced semiconductor to metal transition in the two-dimensional honeycomb structure of MoS_2. Nano Res. **5**(1), 43–48 (2011)
16. Q.-K. Xue, Nanoelectronics: a topological twist for transistors. Nat. Nanotechnol. **6**(4), 197–198 (2011)
17. F. Schwierz, Graphene transistors. Nat. Nanotechnol. **5**(7), 487–496 (2010)
18. M.C. Lemme, T.J. Echtermeyer, M. Baus, H. Kurz, A graphene field-effect device. IEEE Electron Dev. Lett. **28**, 282–284 (2007)
19. X. Li et al., Large-area synthesis of high-quality and uniform Graphene films on copper foils. Science **324**, 1312–1314 (2009)
20. Y.M. Lin et al., 100-GHz transistors from wafer-scale epitaxial graphene. Science **327**, 662 (2010)
21. J.S. Moon et al., Epitaxial-graphene RF field-effect transistors on Si-face 6H-SiC substrates. IEEE Electron. Dev. Lett. **30**, 650–652 (2009)
22. Y. Yang, R. Murali, Impact of size effect on graphene nanoribbon transport. IEEE Electron. Device Lett. **31**(3), 2009–2011 (2010)
23. M. Evaldsson, I. Zozoulenko, H. Xu, T. Heinzel, Edge-disorder-induced Anderson localization and conduction gap in Graphene nanoribbons. Phys. Rev. B **78**(16), 1–4 (2008)
24. B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, A. Kis, Single-layer MoS_2 transistors. Nature Nanotechnol. **6**(3), 147–150 (2011)
25. B. Radisavljevic, M.B. Whitwick, A. Kis, Integrated circuits and logic operations based on single-layer MoS_2. ACS Nano **5**(12), 9934–9938 (2011)
26. R. Grau-Crespo, R. Lopez-Cordero, Comment on 'Ab initio study of MoS_2 and Li adsorbed on the (1010) face of MoS_2' by V. Alexiev, R. Prins and Th. Weber, Phys. Chem. Chem. Phys., 2000, 2, 1815, and 'DFT study of MoS_2 and hydrogen adsorbed on the (1010) face of MoS_2' by V. Alexiev, R. Pri. Phys. Chem. Chem. Phys. **4**(16), 4078–4079 (2002)
27. R.F. Frindt, Single crystals of MoS_2 several molecular layers thick. J. Appl. Phys. **37**(4), 1928 (1966)
28. P. Joensen, R.F. Frindt, S.R. Morrison, Single-layer MoS_2. Mater. Res. Bull. **21**(4), 457–461 (1986)
29. M.A. Gee, R.F. Frindt, S.R. Morrison, Inclusion compounds of MoS_2. Mater. Res. Bull. **21** ((c)), 543–549 (1986)
30. S.R. Morrison, High activity catalyst from exfoliated MoS_2. J. Catal. **345**, 334–345 (1987)
31. J.S. Zabinski, M.S. Donley, V.J. Dyhouse, N.T. McDevitt, Chemical and tribological characterization of PbO-MoS_2 films grown by pulsed laser deposition. Thin Solid Films **214** (2), 156–163 (1992)
32. T. Li, G. Galli, Electronic properties of MoS_2 nanoparticles. J. Phys. Chem. **12**, 16192–16196 (2007)
33. A. Splendiani et al., Emerging photoluminescence in monolayer MoS_2. Nano Lett. **10**(4), 1271–1275 (2010)
34. K.F. Mak, C. Lee, J. Hone, J. Shan, T.F. Heinz, Atomically thin MoS_2: a new direct-gap semiconductor. Phys. Rev. Lett. **105**(13), 136805 (2010)

35. T. Korn, S. Heydrich, M. Hirmer, J. Schmutzler, C. Schüller, Low-temperature photocarrier dynamics in monolayer MoS_2. Appl. Phys. Lett. **102109**, 3 (2011)
36. C. Lee, H. Yan, L.E. Brus, T.F. Heinz, J. Hone, S. Ryu, Anomalous lattice vibrations of single-and few-layer MoS_2. ACS Nano **4**(5), 2695–2700 (2010)
37. A. Ayari, E. Cobas, O. Ogundadegbe, M.S. Fuhrer, Realization and electrical characterization of ultrathin crystals of layered transition-metal dichalcogenides. J. Appl. Phys. **101**(1), 014507 (2007)
38. J. Brivio, D.T.L. Alexander, A. Kis, Ripples and layers in ultrathin MoS_2 membranes. Nano Lett. **11**(12), 5148–5153 (2011)
39. P. Johari, V.B. Shenoy, Tunable dielectric properties of transition metal dichalcogenides. ACS Nano **5**(7), 5903–5908 (2011)
40. J. Sun et al., Spatially-resolved structure and electronic properties of graphene on. ACS Nano **4**(12), 7073–7077 (2010)
41. H. Liu, P.D. Ye, MoS_2 dual-gate MOSFET with atomic-layer-deposited Al_2O_3 as top-gate dielectric. IEEE Electron Device Lett. **33**(4), 546–548 (2012)
42. Y. Zhang, J. Ye, Y. Matsuhashi, Y. Iwasa, Ambipolar MoS_2 thin flake transistors. Nano Lett. **12**(3), 1136–1140 (2012)
43. Y. Yoon, K. Ganapathi, S. Salahuddin, How good can monolayer MoS_2 transistors be? Nano Lett. **11**(9), 3768–3773 (2011)
44. L. Liu, S.B. Kumar, Y. Ouyang, J. Guo, Performance limits of monolayer transition metal dichalcogenide transistors. IEEE Trans. Electron Devices **58**(9), 3042–3047 (2011)
45. S. Ghatak, A.N. Pal, A. Ghosh, Nature of electronic states in atomically thin MOS field-effect transistors. ACS Nano **5**(10), 7707–7712 (2011)
46. D.M.D.J. Singh, T. Pradeep, J. Bhattacharjee, U.V. Waghmare, Closed-cage clusters in the gaseous and condensed phases derived from sonochemically synthesized MoS_2 nanoflakes. Phys. Chem. A **109**, 7339 (2005)
47. S. Bhattacharyya, A.K. Singh, Semiconductor-metal transition in semiconducting bilayer sheets of transition metal dichalcogenides. arXiv:1203.6820v1 [cond-mat.mtrl-sci] 30 Mar 2012

Quantum Nanoengineering

Arkaprava Bhattacharyya and Vijay K. Arora

1 Particle in a Box

Particle in a box is a famous Quantum Mechanical problem that can be solved analytically. Although very simple, this model provides understanding of QM by asserting discrete energy levels to the particle trapped inside a narrow well. Figure 1 represents the nature of the well where the particle is constrained to move in one dimension. The particle does not experience any force inside the well but cannot cross the barrier as the barrier is having an infinite potential. The potential profile is given by,

$$\begin{aligned} &V = 0, \quad 0 \le x \le L \\ &V = \infty, \quad x < 0,\ x > L \end{aligned} \tag{1}$$

To analyze the wave character of an electron and corresponding energy levels, it is necessary to solve the time-independent Schrödinger equation inside the well.

A. Bhattacharyya (✉) · V.K. Arora
Department of Electronic and Computer Engineering, Faculty of Electrical Engineering,
Universiti Teknologi Malaysia UTM, 81310 Skudai, Johor, Malaysia
e-mail: arkapravabhattacharyya@eee.sastra.edu

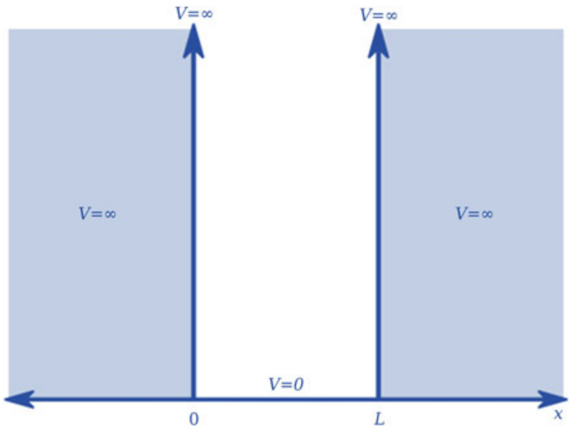
V.K. Arora
e-mail: vijay.arora@wilkes.edu

A. Bhattacharyya
School of Electrical and Electronics Engineering, SASTRA University,
Tirumalaisamudram, Thanjavur 613401, Tamil Nadu, India

V.K. Arora
Department of Electrical Engineering and Physics, Wilkes University,
84 West South Street, Wilkes-Barre, PA 18766, USA

© Springer-Verlag Berlin Heidelberg 2015
A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_4

Fig. 1 Infinite potential well in 1 Dimension



$$\frac{d^2\psi}{dx^2}+\frac{2m^*}{\hbar^2}E\psi=0 \tag{2}$$

Equation (2) simplifies to

$$\frac{d^2\psi}{dx^2}+k^2\psi=0,\quad E=\frac{\hbar^2k^2}{2m} \tag{3}$$

where m is the mass of the particle inside the well. As the particle is surely inside the box, we can apply a boundary condition $\psi(0)=\psi(L)=0$. Now, the non-trivial solution of Eq. (3) is simple:

$$\psi(x)=A\sin(kx),\quad k=\frac{n\pi}{L},\ n=1,2,3\ldots \tag{4}$$

The quantum condition in Eq. (4) when substituted in the energy $E=\hbar^2k^2/2m$ expression of Eq. (3) gives the quantized energy levels. Applying the normalization condition, one can extract the value of constant A.

$$\int_0^L \psi*\psi dx=1\Rightarrow A=\sqrt{\frac{2}{L}} \tag{5}$$

Including the value of A, the normalized wavefunctions (also known as eigenfunctions) and quantized energy values (also known as eigenvalues) are then given by

$$\psi(x)=\sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x}{L}\right),\quad E_n=n^2\frac{\pi^2\hbar^2}{2mL^2}=n^2E_1 \tag{6}$$

Assuming an electron inside the box, Fig. 2 gives the quantized wavefunctions in (a) and the description of the energy levels in (b) for a quantum well of 10 nm length ($L_x = 10\,\mathrm{nm}$). The energy spacing increases as energy moves higher. On the other hand, the levels tend to come closer together in a hydrogen atom. The spectrum indicated is thus inverted spectrum of that of the hydrogen level.

In case of a 3D box, electron wavefunctions are obtained by multiplying independent contributions coming from each dimension. For a 3D box, the complete time-independent solution and energy eigenvalues look like,

$$\psi(x, y, z) = \sqrt{\frac{8}{L_x L_y L_z}} \sin\left(\frac{n_x \pi x}{L_x}\right) \sin\left(\frac{n_y \pi y}{L_y}\right) \sin\left(\frac{n_z \pi z}{L_z}\right) \tag{7}$$

$$E_n = \frac{\pi^2 \hbar^2}{2m} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2}\right) \tag{8}$$

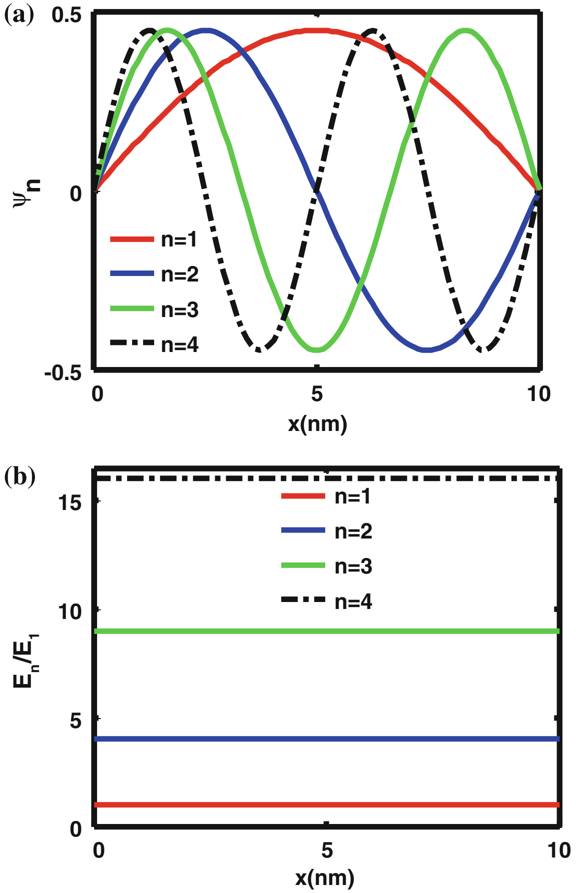


Fig. 2 **a** Electron wavefunctions in a quantum well of 10 nm length, **b** Normalized discrete energy levels inside a quantum well

The quantum analysis of particle in a box system is readily applicable to electrons inside a semiconducting material or in a heterostructure device. In the case of electron trapped in a 3D quantum well, electron wavefunctions and energy eigenvalues are given by Eqs. (7) and (8). We should keep in mind that electrons inside a semiconductor should be modeled using their "effective mass" not by using electron rest mass. Effective mass of electron is calculated from Density of States function.

2 Quantum Limit: From 3D to 0D

Dimensionality of a quantum device is obtained by comparing its size in each coordinate with the De Broglie wavelength (λ_F) at the Fermi energy (E_F). Near the bottom of the conduction band in a semiconductor, the electron relation becomes $E_F = \frac{\hbar^2 k^2}{2m_e^*}$. Corresponding De Broglie wavelength is

$$\lambda_F = \frac{h}{p} = \frac{h}{\sqrt{2m_e^* E_F}} \tag{9}$$

Here, m_e^* denotes electron effective mass inside a semiconductor. Considering a 3D structure, λ_F becomes [2],

$$\lambda_F = \frac{2\pi}{\left(3N_e\pi^2\right)^{1/3}} \tag{10}$$

where, N_e is the effective density of states. Computing the values of λ_F for different metals and semiconductors, it comes out

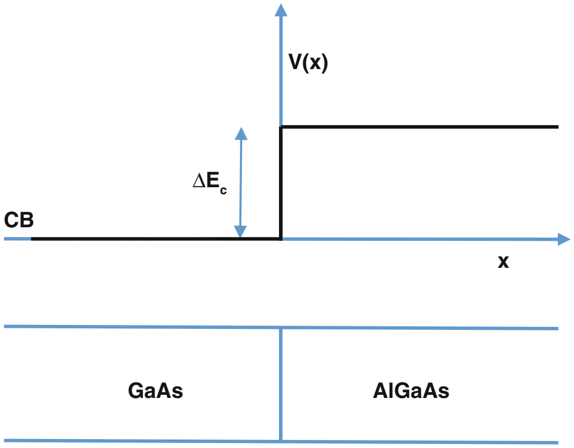
$$\lambda_F \sim 0.5 - 1\,\text{nm} \quad \text{for metals}$$
$$\lambda_F \sim 10 - 100\,\text{nm} \quad \text{for semiconductors}$$

If any of the dimensions of a device is below the specified limit (λ_F), that dimension becomes quantized (or digital). In a bulk semiconductor, no dimension is quantized (all analog no digital). In a 2D quantum well, one dimension is quantized but other two dimensions are analog. In a quantum wire, two dimensions are quantized and only one dimension is allowed to carry analog energy levels. In a quantum dot, all dimensions are quantized making it a quantum 0D system.

2.1 *Quantum Confinement in Semiconductors*

A potential step or quantum well can be formed in heterostructure devices where different parts consisting the device are having different bandgaps. For example, GaAs

Fig. 3 A simple illustration of potential stepformed at the jucntion of GaAs and AlGaAs. Only the Conduction Band (CB) is shown

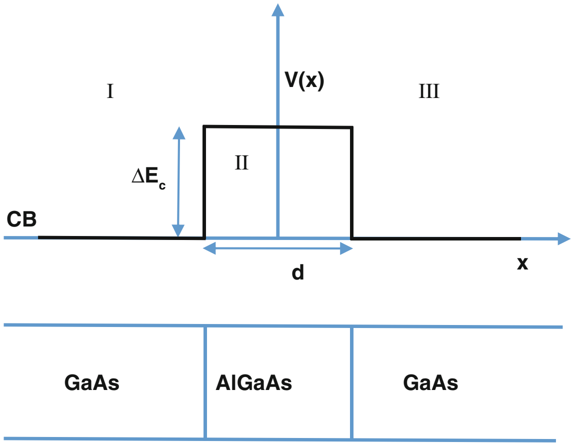


layer is combined with AlGaAs layer to make a potential step (Fig. 3) and similarly a quantum well can be formed by sandwiching GaAs layer in between two layers of AlGaAs (Fig. 4). AlAs has a larger bandgap (2.16 eV) than that of GaAs (1.42 eV). By alloying AlAs with GaAs, it is possible to vary the bandgap depending on the percentage x of AlAs to make a layer of $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$. The fabrication of this alloy is attractive due to the almost perfect match of the lattice constants of AlAs and GaAs that make up this alloy.

2.1.1 Potential Step

Here we consider GaAs layer in conjugation with AlGaAs to form a potential step as shown in Fig. 3. The conduction band (CB) discontinuity occurs due to the difference of band gap between GaAs and AlGaAs.

Fig. 4 A simple illustration of potential barrier showing the barrier height (ΔE_c) and barrier width (d). Three different regions for electron wave are also shown



If the conduction band electron energy is below the step potential ($0 < E < \Delta E_c$), classically the electron cannot pass but quantum mechanically there is a finite probability that the electron can tunnel through the potential barrier with a decaying exponential function.

2.1.2 Potential Barrier

Figure 4 shows a 1D potential barrier which is formed by adding another layer of GaAs to the potential step (Fig. 3). Quantum mechanically, electrons coming from the left-hand side with en energy less than the barrier height ($0 < E < \Delta E_c$), can 'tunnel' trough the barrier.

Figure 4 shows the propagation of the electron waves reflecting in Region I, decaying in Region II, and again propagating in Region III. Electron wavefunctions are obtained by solving Schrödinger equation for different regions.

$$\begin{array}{lll} \text{Region I: } \psi_1(x) = A\mathrm{e}^{jkx} + B\mathrm{e}^{-jkx} & \Rightarrow & k = \sqrt{\frac{2m_e^* E}{\hbar^2}} \\ \text{Region II: } \psi_2(x) = C\mathrm{e}^{-\alpha x} + D\mathrm{e}^{+\alpha x} & \Rightarrow & \alpha = \sqrt{\frac{2m_e^*(\Delta E_c - E)}{\hbar^2}} \\ \text{Region III: } \psi_3(x) = F\mathrm{e}^{jkx} & \Rightarrow & k = \sqrt{\frac{2m_e^* E}{\hbar^2}} \end{array} \tag{11}$$

Wavefunctions are assumed to be continuous and differentiable at the boundaries. That provides us valuable boundary conditions to determine coefficients in Eq. (11).

$$\begin{array}{lll} \psi_1(x)|_{x=0} = \psi_2(x)|_{x=0} & \text{and} & \left.\frac{\mathrm{d}\psi_1(x)}{\mathrm{d}x}\right|_{x=0} = \left.\frac{\mathrm{d}\psi_2(x)}{\mathrm{d}x}\right|_{x=0} \\ \psi_2(x)|_{x=d} = \psi_3(x)|_{x=d} & \text{and} & \left.\frac{\mathrm{d}\psi_2(x)}{\mathrm{d}x}\right|_{x=d} = \left.\frac{\mathrm{d}\psi_3(x)}{\mathrm{d}x}\right|_{x=d} \end{array} \tag{12}$$

Applying boundary conditions to solve the Schrödinger equation for a rectangular barrier, the tunneling coefficient becomes [3],

$$T = \left|\frac{F}{A}\right|^2 = \frac{4}{4\cos h^2(\alpha d) + \left[\frac{\alpha}{k} - \frac{k}{\alpha}\right]^2 \sin h^2(\alpha d)} \qquad E < \Delta E_c \tag{13}$$

where d is the barrier depth.

2.1.3 Quantum Well

The prototype quantum well (Fig. 5) is often modeled with the one with infinite boundaries as the band-gap discontinuity ΔE_c (the fraction of the band-gap

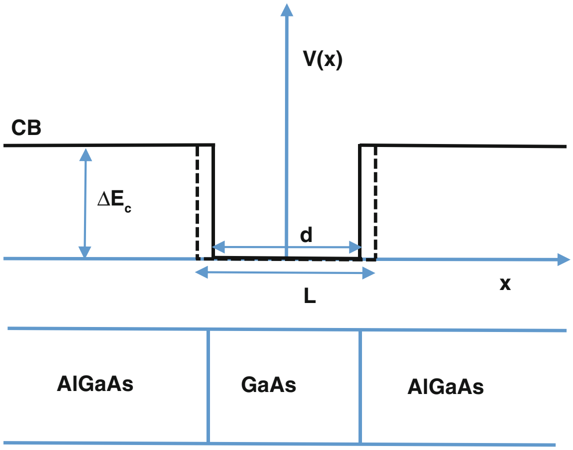


Fig. 5 A prototype AlGaAs/GaAs/AlGaAs quantum well where GaAs (bandgap E_g = 1.43 eV) is sandwiched between two wide bandgap layers of AlGaAs layers. The physical width of the quantum well is denoted by *d* and the effective width is shown by *L* letter

difference in the conduction edges) is much larger than the thermal energy k_BT. As some parts of the wavefunction penetrates into the regions of AlGaAs, the effective width *L* becomes larger than the physical width *d*. In that case, *L*, in terms *d* is given by [1].

$$L = \left(1 + \frac{1}{P}\right)d, \quad P = \frac{d}{2}\sqrt{\frac{2m_e^* \Delta E_c}{\hbar^2}} \tag{14}$$

where *P* is the strength of the quantum well. *P* defines the extent to which wavefunctions penetrate in the classically forbidden region, $P \to \infty$ being the ideal. More description on GaAs quantum well is available in [4].

2.2 3D Density of States

In the previous example, we have calculated particle wavefunctions and corresponding energy values. We also want to know how many states are available for the particle to fill within an energy range. Assuming the particle as an electron trapped inside an electronic device, the total number of available states per unit area per unit energy is termed as density of states of electron in the device. In this section, we focus on calculating Density of States (DOS) of a 3D material.

As shown in Eq. (4), $k_{x,y,z} = n_{x,y,z}\pi / L_{x,y,z}$ are the digitized momenta in all directions when the confinement in all three directions is considered. The energy is the same for $\pm k_{x,y,z}$ as it depends on the square of $k_{x,y,z}$. Considering both positive and negative $k_{x,y,z}$'s, the change in momentum becomes $\Delta k_{x,y,z} = 2\Delta n_{x,y,z}\pi / L_{x,y,z}$ with a factor of 2 added. In other words, total number of quantum states possible for the trapped particle contained are [5]

$$\Delta n_x \Delta n_y \Delta n_z = \frac{L_x L_y L_z}{(2\pi)^3} \Delta k_x \Delta k_y \Delta k_z \tag{15}$$

Since energy $E = \frac{\hbar^2 k^2}{2m_e^*}$ depends only on the magnitude of k with $k^2 = k_x^2 + k_y^2 + k_z^2$, the use of spherical polar coordinates is appropriate. In that system, $\Delta k_x \Delta k_y \Delta k_z = k^2 \mathrm{d}k \, \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$. Integrating over $\theta = 0 \rightarrow \pi$ and $\phi = 0 \rightarrow 2\pi$ covers the complete energy sphere with radius k ink-space. Inserting the integration value in Eq. (15) gives total number of available states per unit volume.

$$\frac{\Delta n_x \Delta n_y \Delta n_z}{L_x L_y L_z} = 2\frac{1}{(2\pi)^3} k^2 \, \mathrm{d}k \int\limits_0^{\pi} \sin\theta \, \mathrm{d}\theta \int\limits_0^{2\pi} \mathrm{d}\phi = 2\frac{4\pi}{(2\pi)^3} k^2 \, \mathrm{d}k \tag{16}$$

where factor 2 on the left-hand side accounts for 2 spin states as each state is occupied by two electrons (spin up and spin down). Using $k = \left(\frac{2m_e^*}{\hbar^2}\right)^{1/2} E^{1/2}$, the density of states (DOS), $D(E)$ (total no of states per unit volume per unit energy), is given by

$$\begin{aligned} D_{3e}(E) &\equiv \frac{\Delta n_x \Delta n_y \Delta n_z}{L_x L_y L_z dE} = 2\frac{4\pi}{(2\pi)^3} \left(\frac{2m_e^*}{\hbar^2}\right)^{3/2} \frac{1}{2} E^{1/2} \\ &= \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2}\right)^{3/2} E^{1/2} \end{aligned} \tag{17}$$

If the electron's potential energy is E_{co} (conduction band minima), the DOS becomes [5],

$$D_{3e}(E) = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2}\right)^{3/2} (E - E_{\mathrm{co}})^{1/2} \tag{18}$$

The DOS for the holes in the valence band can be similarly derived

$$D_{3h}(E) = \frac{1}{2\pi^2} \left(\frac{2m_h^*}{\hbar^2}\right)^{3/2} (E_{\mathrm{vo}} - E)^{1/2} \tag{19}$$

$D_{3e}(E)$ for the free electron mass ($m_e^* = m_o$) is given by

$$D_3(E) = 1.06 \times 10^{56}\,\mathrm{J}^{-3/2}\,\mathrm{m}^{-3}(E_{\mathrm{Joules}} - E_{\mathrm{co}})^{1/2} \tag{20}$$

In terms of conventional units of eV for energy and m^{-3} for volume, Eq. (18) transforms for a semiconductor with an effective mass m^* to [5]

$$D_3(E) = 6.8 \times 10^{27}\,\mathrm{eV}^{-3/2}\,\mathrm{m}^{-3}\left(\frac{m^*}{m_o}\right)^{3/2}\left(E_{\mathrm{eV}} - E_{\mathrm{co,eV}}\right)^{1/2} \tag{21}$$

This derivation demonstrates how analog 3D world is derived from the digital world where spacing between the energy levels is very small because of the large size in each of the three dimensions. The quantum energy levels are packed closely together and their density per unit energy per unit volume is what appears in Eq. (17).

2.3 2D Model

In a quasi-two-dimensional (Q2D) quantum well, one direction (say z-direction), for example, is below the De Broglie wavelength ($L_z < \lambda_F$) while the other two $L_{x,y} \gg \lambda_F$ maintain their bulky character. In this Q2D quantum well, analog-type levels appear only in two dimensions while third dimension is quantized.

2.3.1 Energy Eigenvalues

The total energy for an electron is the sum of energies of one quantized and two unquantized levels [5].

$$E_k = E_{\mathrm{co}} + \frac{\hbar^2(k_x^2 + k_y^2)}{2\,m_e^*} + n_z^2 E_{1z} = E_c + \frac{\hbar^2(k_x^2 + k_y^2)}{2\,m_e^*} \tag{22}$$

with

$$E_c = E_{\mathrm{co}} + n_z^2 E_{1z}, \quad n_z = 1, 2, 3, \ldots$$

In the extreme case ($n_z = 1$) the spacing between the levels in the z-direction is so large that it is impossible for large states to be populated, $E_c = E_{\mathrm{co}} + E_{1z}$. Q2D layer becomes truly 2D in this extreme. The conduction band is lifted by zero-point energy E_{1z} (minimum energy in a quantum wave). Similarly, the valence band $E_v = E_{\mathrm{vo}} - E_{\mathrm{1hz}}$ drops by zero-point energy E_{1hz} of holes.

2.3.2 Density of States (DOS)

Following the same pattern as before (with spin 2 factor) and noting that $\Delta k_x \Delta k_y = kdkd\theta$ in circular coordinates with $\theta = 0 \rightarrow 2\pi$ (going full circle), the DOS $D_{2e(h)}$ for electrons e (or for holes h) is given by [5]

$$D_{2e(h)}(E) \equiv \frac{\Delta n_x \Delta n_y}{L_x L_y dE} = \frac{2\pi}{(2\pi)^2}\left(\frac{2m^*_{e(h)}}{\hbar^2}\right) = \frac{m^*_{e(h)}}{\pi\hbar^2} \tag{23}$$

DOS in this case are the number of quantum states per unit area per unit energy. These can be numerically shown to lead to

$$D_{2e(h)}(E) = \frac{m^*_{e(h)}}{m_o} \times 2.6 \times 10^{37}\,\mathrm{J}^{-1}\,\mathrm{m}^{-2} = \frac{m^*_{e(h)}}{m_o} \times 4.17 \times 10^{18}\,\mathrm{eV}^{-1}\,\mathrm{m}^{-2} \tag{24}$$

2.4 1D Model

As two (say y- and z-direction) are squeezed so that $L_y \text{ and } L_z < \lambda_F$ while the remaining one direction $L_x \gg \lambda_F$ maintains its analog character, a quasi-one-dimensional (Q1D) quantum well or nanowire emerges.

2.4.1 Energy Eigenvalues

In this Q1D quantum well, analog-type levels appear only in one dimension while the other two go quantum:

$$E_k = E_{\mathrm{co}} + \frac{\hbar^2\,k_x^2}{2\,m_e^*} + n_y^2 E_{1y} + n_z^2 E_{1z}, \quad n_{y,z} = 1, 2, 3, \ldots \tag{25}$$

The conduction band is now lifted by the quantized levels and becomes, $E_c = E_{\mathrm{co}} + n_y^2\,E_{1y} + n_z^2\,E_{1z}$ giving a simpler $E{-}K$ relation [5],

$$E_{nk} = E_c + \frac{\hbar^2\,k_x^2}{2\,m_e^*} \tag{26}$$

In the extreme case ($n_{y,z} = 1$) the spacing between the levels in the y- and z-direction is so large that it is impossible for large states to be populated, $E_c = E_{\mathrm{co}} + E_{1y} + E_{1z}$. In this extreme, Q1D layer becomes truly 1D (a nanowire). The conduction band is lifted by what is known as zero-point energy of electrons

$E_{1y} + E_{1z}$. Similarly, the valence band $E_v = E_{\mathrm{vo}} - E_{1\mathrm{hy}} - E_{1\mathrm{hz}}$ drops by zero-point energy of the holes $E_{1\mathrm{hy}} + E_{1\mathrm{hz}}$ [5].

2.4.2 Density of States (DOS)

Following the same pattern as before (with spin 2 factor) and noting that $\Delta k_x = 2k_x$ (positive and negative direction), the DOS $D_{1e(h)}$ (E) for electrons and for holes is given by

$$\begin{aligned} D_{1e}(E) &\equiv \frac{\Delta n_x}{L_x dE} = \frac{1}{\pi}\left(\frac{2m_e^*}{\hbar^2}\right)^{1/2}(E - E_c)^{-1/2} \\ D_{1h}(E) &\equiv \frac{\Delta n_x}{L_x dE} = \frac{1}{\pi}\left(\frac{2m_h^*}{\hbar^2}\right)^{1/2}(E_h - E)^{-1/2} \end{aligned} \tag{27}$$

The factor of 2 is due to spin and the other factor of 2 is for $\pm k_x$. Numerically, for a semiconductor, it can be put in the form [5]

$$\begin{aligned} D_1(E) &= 4.07 \times 10^{18}\,\mathrm{J}^{-1/2}\,\mathrm{m}^{-1}\left(\frac{m^*}{m_o}\right)^{1/2}(E_J - E_c)^{-1/2} \\ &= 1.63 \times 10^9\left(\frac{m^*}{m_o}\right)^{1/2}\,\mathrm{eV}^{-1/2}\,\mathrm{m}^{-1}(E_{\mathrm{eV}} - E_c)^{-1/2} \end{aligned} \tag{28}$$

Figure 6 shows the comparison between DOS in 3D, 2D, and 1D for GaAs ($m^*/m_o = 0.067$) and for a quantum well with $L_z = 10$ nm (2D) and $L_y = L_z = 10$ nm (1D). These curves are normalized so density of states is in common unit $\mathrm{eV}^{-1}\,\mathrm{nm}^{-3}$.

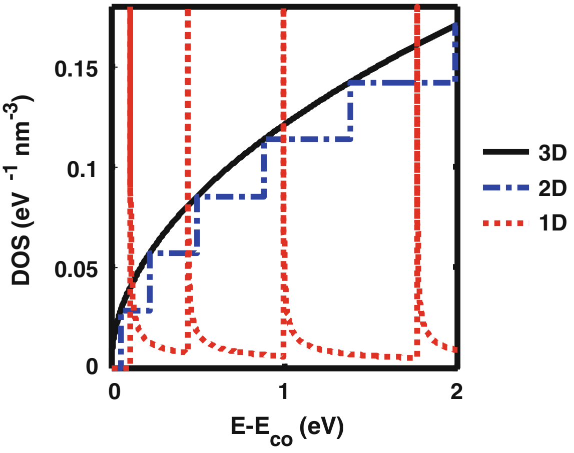


Fig. 6 Comparison between the DOS for 3D, 2D, and 1D for GaAs ($m^*/m_o = 0.067$) and for a quantum well with $L_z = 10$ nm (2D) and $L_y = L_z = 10$ nm (1D) [5]

2.5 Q0D Model

The DOS concept is not valid in a quantum dot as there is no densely packed quantum levels. In a quantum dot all three Cartesian directions are squeezed so that L_x, L_y and $L_z < \lambda_F$. Quantum dot or box emerges without any analog-like energy levels.

2.5.1 Energy Eigenvalues

The quantized energy is given by:

$$E_{n_x,n_y,n_z} = E_{\mathrm{co}} + n_z^2\, E_{1z} + n_y^2\, E_{1y} + n_x^2\, E_{1x}, \quad n_{x,y,z} = 1, 2, 3, \ldots \tag{29}$$

For a spherical dot of radius R, the wavefunctions are spherical Bessel function. However, in its approximate form, it can be obtained as

$$E_n = n^2 \frac{\pi^2 \hbar^2}{2m_e^* R^2} = n^2 E_1 \tag{30}$$

For an R = 6.2 nm radius sphere, which has the same volume as 10-nm cube, $\Delta E = E_2 - E_1 = 0.438$ eV for GaAs. On the other hand R = 10 nm can be equated to L = 16.1 nm for which $\Delta E = E_2 - E_1 = 0.169$ eV.

For a particle bound in a spherical well with infinite boundaries, the energy levels are given by [5]

$$E_n = \alpha_{n,\ell}^2 \frac{\hbar^2}{2m * R^2} = n^2 E_1 \tag{31}$$

where $\alpha_{1,(0-4)} = 3.142,\ 4.493,\ 5.763,\ 6.988, \text{ and } 8.183$ and $\alpha_{2,(0-4)} = 6.283,\ 7.725, 9.095, 10.417, \text{ and } 11.705.$ $\alpha_{1,0} \approx \pi.$ The wave functions become, $\psi_{n,\ell,m}(r,\ \theta,\ \phi) = A j_n(\alpha_{n,\ell} r) Y_\ell^m(\theta, \phi)$ where $j_n(\alpha_{n,\ell} r)$ is the Bessel function with $j_n(\alpha_{n,\ell} R) = 0$ and $Y_\ell^m(\theta, \phi)$ are spherical harmonics for spherical symmetric system [5].

2.6 Conclusion

As the device dimensions are downscaled to handle "More than Moore" phase, application of quantum mechanics has become necessary. Nowadays, different band-gap materials with a similar lattice constant are combined to form a heterostructure which are essential for achieving high speed carrier transport. This chapter provides a foundation to quantum assumptions that are directly applicable to lower dimensional heterostructure devices. Some simple illustrations are shown to describe quantum tunneling phenomena. It is now clear that lower dimensional devices behave differently than the bulk 3D device and one needs to calculate the Density of States (DOS) function separately for quasi 2D or 1D devices. The DOS function and corresponding energy eigenvalues are presented in a concise manner for low dimensional devices. Quantum well (QW), on the other hand, has shown applications especially in the field of optoelectronics as QW laser diodes are popular due to their capability of producing really short wavelengths.

Overall, in this chapter, a basic treatment of quantum devices is presented to serve a starting point which can be further furnished to get a deeper understanding.

Acknowledgments VKA thanks the Universiti Teknologi Malaysia (UTM) for an award of distinguished visiting professorship and UTM Research University Grant (GUP) Q.J130000.2623.04H32 of the Ministry of Education (MoE).

References

1. A.T.M. Fairus, V.K. Arora, Quantum engineering of nanoelectronic devices: the role of quantum confinement on mobility degradation. Microelectron. J. **32**, 679–686 (2001)
2. G.W. Hanson, *Fundamental of Nanoelectronics* (Pearson/Prentice-Hall, Upper Saddle River, NJ, 2008)
3. J. Singh, *Semiconductor Devices: An Introduction* (McGraw-Hill College, NY, 1994)
4. B.G. Streetman, S.K. Banerjee, *Solid State Electronic Devices*, 6th edn. (Prentice-Hall, Upper Saddle River, 2006)
5. V.K. Arora, *Nanoelectronics: Quantum Engineering of Low-Dimensional Nanoensembles* (CRC Press, Boca Raton, FL, 2015)

Quantum Effect on Properties of Nanomaterials

Chandan Kumar Ghosh

1 Introduction

The physical properties of the material significantly depend on their size when reduced to nanodimension. In bulk system, most of the atoms are the interior atoms, so the properties of the materials are determined by them. But in the case of nanoparticles, the fraction of surface atoms cannot be neglected. To illustrate the increasing contribution of surface atoms, let us consider the following example of spherical nanoparticle having radius "r." Thus surface-(S)-to-volume (V) ratio $(S/V) = 3/r$ for such nanoparticle dramatically increases as its size is reduced Since surface atoms have different environment in contrast to interior atoms, their behaviors to any external excitation are completely different. The size effect becomes prominent when their size is $\sim$5–50 nm. In this chapter, we have discussed the effect of size on few properties such as melting point, optical property, magnetic property, and mechanical property of materials. Section 1 describes the size effect on the melting point of nanoparticles, where it will be shown that the melting point of an isolated particle decreases as its size is reduced. In addition to that, the developed theory also predicts that the melting point can also be different for different morphologies of the nanoparticle. Nanoparticles, embedded within matrix, exhibit an enhancement of the melting point. This phenomenon is known as superheating of nanoparticles. We will illustrate the shape and size effect on the electronic structure and related optical properties of the nanomaterials in Sect. 2. The most striking effect of dimension reduction is that the energy dependence of the density of states (DOS) exhibits different functional form. The dependence may generalized in the form $D(E) \sim E^{(d/2-1)}$, where "d" represents the dimension of the system. Experimentally, band gap of the material was found to be increased with

---

C.K. Ghosh (✉)
School of Materials Science and Nanotechnology, Jadavpur University,
Kolkata 700 032, India
e-mail: chandu_ju@yahoo.co.in

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_5

reduction of size. Quantum confinement effect was taken into account to explain band gap enhancement. The other important optical property that has been considered here is the excitonic effect. Instead of moving independently, electron in conduction band and hole in valence band form a stable quasiparticle state that is known as exciton. In this section, we have discussed the confinement effect on the excitonic property of the materials. In the following section, we have considered the magnetic property of the materials. Among different magnetic materials, developed so far, ferromagnetic materials have drawn special attraction due to their huge applicational opportunity in the field of magnetic storage devices, drug delivery, etc. The size effect on magnetic materials, mostly ferromagnetic materials, has been discussed. Normally ferromagnetic materials are characterized by saturation magnetization, coercive field, and remanent magnetization. In this section, we have elaborately discussed the size effect on these properties of the ferromagnetic materials. Most importantly, there exists a characteristic dimension of the ferromagnetic materials, below which they possess paramagnetic property. In next two sections, we have given very short overview of the effect of size on the mechanical and dielectric properties.

2 Melting Point

Study of thermodynamic properties of nanomaterials becomes an interesting area of research because they are being widely used in many fields such as microelectronics, solar energy, nonlinear optics, magnetic storage devices, etc. For better usage of nanomaterials, it is necessary to know their stability. In this context, knowledge of melting point is supposed to illustrate some idea about the stability of nanomaterials. Melting point is defined as the temperature at which the solid transforms into liquid at a fixed pressure.

It was believed long back that the melting point of any material is an intrinsic property of it. But in recent time, it has been well established that the melting point, in general, decreases as its size approaches to nanoscale. The origin of such depression of meltingpoint on the size, shape of the nanomaterial has been a long mystery that attracted researchers over the years and enormous theoretical and experimental research work had been carried out by scientists during recent past. After vast research, it is now established that the changes in the melting point of the solid particles occurs in the nanoscale material due to large surface-to-volume ratio compare to its bulk counterpart. Since surface-to-volume ratio depends on the shape of the particles, therefore this consideration is also be able to the shape dependency of the melting point and subsequently other thermodynamical properties of solid materials.

In general, melting temperature is assumed to be proportional to the cohesive energy of the material, i.e., the thermal energy required to divide the material into their individual constituting elements. There exist different methods such as density functional theory, molecular dynamics, cellular method, etc., to calculate cohesive energy of the bulk system. Thus following these methods, it becomes possible to calculate melting point of nanoparticles. Alternatively, few simple methods also exist to simulate the cohesive energy of the nanoparticles taking into account size and shape dependency of the coordination number of atoms, surface packing factors, etc. Though these methods cannot predict melting point very accurately, but the variation of melting point can be well established by them. One of these methods is the consideration of cohesive energy ($E_{\mathrm{c,bulk}}$) of a bulk particle as the sum of total bond energy, i.e., we may write [1]

$$E_{\mathrm{c,bulk}} = \frac{1}{2} n_{\mathrm{t}} \beta_{\mathrm{L}} \varepsilon \tag{1}$$

where n_{t}, β_{L} and ε represent total number of atoms within the system, coordination number of atoms in the solids (number of bonds made by atoms), and energy of each bonds, respectively. Half is introduced to avoid the double counting. From this expression, it may easily be concluded that material with higher bonding energy possesses higher melting point, i.e., higher stability. Atoms, in the interior region of the particle, are strongly bounded with their neighbor atoms. But surface atoms are less tightly bounded to the particle since they have less numbers of neighboring atoms. In bulk systems, contribution of surface atoms to the total cohesive energy is neglected because their contribution to the total energy is negligible small (due to very small surface-to-volume ratio). But for an isolated single nanoparticle having nonnegligible fraction of surface atom compare to total number of atoms, we have to consider their contribution also, hence their cohesive energy ($E_{\mathrm{c,nano}}$) must be written as,

$$E_{\mathrm{c,nano}} = \frac{1}{2} n_{\mathrm{i}} \beta_{\mathrm{L}} \varepsilon + \frac{1}{2} n_{\mathrm{s}} \beta_{\mathrm{S}} \varepsilon \tag{2}$$

where n_{i}, n_{s} represent number of interior and surface atoms, respectively, i.e., total number of atoms in the particles $n_{\mathrm{t}} = n_{\mathrm{i}} + n_{\mathrm{s}}$. β_{S} is the coordination numbers of surface atoms, i.e., the number of bonds made by surface atoms in the surface crystalline planes. Since the melting point is assumed to be proportional to cohesive energy of the particle, so we may write,

$$\begin{aligned} \frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} &= \frac{E_{\mathrm{c,nano}}}{E_{\mathrm{c,bulk}}} = \frac{\frac{1}{2} n_{\mathrm{i}} \beta_{\mathrm{L}} \varepsilon + \frac{1}{2} n_{\mathrm{s}} \beta_{\mathrm{S}} \varepsilon}{\frac{1}{2} n_{\mathrm{t}} \beta_{\mathrm{L}} \varepsilon} = \frac{\frac{1}{2} (n_{\mathrm{t}} - n_{\mathrm{s}}) \beta_{\mathrm{L}} + \frac{1}{2} n_{\mathrm{s}} \beta_{\mathrm{S}}}{\frac{1}{2} n_{\mathrm{t}} \beta_{\mathrm{L}}} \\ &= 1 - \frac{(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})}{\beta_{\mathrm{L}}} \frac{n_{\mathrm{s}}}{n_{\mathrm{t}}} \end{aligned} \tag{3}$$

According to Eq. (3), it is clear that the melting point depends on the fraction of the surface atoms and it will decrease as this fraction increases as in the case of nanoparticles. To calculate the fraction of surface atoms, we define outer surface that consists of many crystalline planes, i.e., the planes which divide the surface atoms into two equal parts. Therefore, volume of the nanoparticles (V_{nano}) may be written as in terms of atomic volume (V_{atom}),

$$n_{\mathrm{i}} V_{\mathrm{atom}} + \frac{1}{2} n_{\mathrm{s}} V_{\mathrm{atom}} = P_{\mathrm{L}} V_{\mathrm{nano}} \tag{4}$$

where P_{L} is defined as the volume packing fraction of the particle. Now Eq. (4) can be re-written as

$$P_{\mathrm{L}} \frac{V_{\mathrm{nano}}}{V_{\mathrm{atom}}} = n_{\mathrm{i}} + \frac{1}{2} n_{\mathrm{s}} = n_{\mathrm{t}} - \frac{1}{2} n_{\mathrm{s}} \tag{5}$$

To define number of surface atoms, surface packing fraction (P_{S}) is defined as the fraction of surface of the nanoparticle occupied by atomic surfaces. So, n_{s} can be written in terms of P_{S} in the following manner,

$$n_{\mathrm{s}} = P_{\mathrm{S}} \frac{4\pi R^2}{\pi r^2} \tag{6}$$

where R and r represent the radius of nanoparticles and atoms, respectively. Combining Eqs. (5) and (6), we may write the following expression for total number of atoms,

$$n_{\mathrm{t}} = P_{\mathrm{L}} \frac{R^3}{r^3} + P_{\mathrm{S}} \frac{2R^2}{r^2} \tag{7}$$

Substituting the expressions for n_{s} and n_{t} from Eqs. (6) and (7) into Eq. (3), it may be written

$$\frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} = 1 - \frac{(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})}{\beta_{\mathrm{L}}} \frac{4P_{\mathrm{S}} \frac{R^2}{r^2}}{P_{\mathrm{L}} \frac{R^3}{r^3} + 2P_{\mathrm{S}} \frac{R^2}{r^2}} \tag{8}$$

In general, $P_{\mathrm{L}} \frac{R^3}{r^3} \gg 2P_{\mathrm{S}} \frac{R^2}{r^2}$, so we may neglect the second term in the denominator of Eq. (8). With this assumption, Eq. (8) may be simplified into the following equation,

$$\frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} = 1 - \frac{(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})}{\beta_{\mathrm{L}}} \frac{4P_{\mathrm{S}} \frac{R^2}{r^2}}{P_{\mathrm{L}} \frac{R^3}{r^3}} = 1 - \frac{(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})}{\beta_{\mathrm{L}}} 4 \frac{P_{\mathrm{S}}}{P_{\mathrm{L}}} \frac{r}{R} \tag{9}$$

In general, $\beta_{\mathrm{L}} > \beta_{\mathrm{S}}$ for any free standing particles, so the second term is always positive. Therefore, we may conclude from Eq. (9) that $T_{\mathrm{mp,nano}} < T_{\mathrm{mp,bulk}}$ and it will decrease monotonically if we decrease the particle size. If we have no idea about the coordination number of surface atoms, then common procedure to analyze the results is that the experimentally measured melting point for various size of nanoparticle is fitted with $q = \frac{\beta_{\mathrm{S}}}{\beta_{\mathrm{L}}}$ and best fitted curve is used for interpretation.

Similar type expression can also be derived for other particles. For this purpose, we consider disk-shaped nanoparticles with radius R and height H. Then for such nanoparticles, the number of surface atoms may be written in terms of surface packing fraction as

$$n_\mathrm{s} = P_\mathrm{S}\frac{2\pi RH + 2\pi R^2}{\pi r^2} = P_\mathrm{S}\frac{2RH + 2R^2}{r^2} \tag{10}$$

and

$$n_\mathrm{i}V_\mathrm{atom} + \frac{1}{2}n_\mathrm{s}V_\mathrm{atom} = P_\mathrm{L}\pi R^2 H \tag{11}$$

From Eqs. (10) and (11), we may write expression for total number of atoms within the system

$$n_\mathrm{t} = P_\mathrm{L}\frac{\pi R^2 H}{\frac{4}{3}\pi r^3} - \frac{1}{2}n_\mathrm{s} = P_\mathrm{L}\frac{\pi R^2 H}{\frac{4}{3}\pi r^3} - P_\mathrm{S}\frac{RH + R^2}{r^2} \tag{12}$$

Now substituting the expression for n_t and n_s into Eq. (3), it is possible to write

$$\frac{T_\mathrm{mp,nano}}{T_\mathrm{mp,bulk}} = 1 - \frac{(\beta_\mathrm{L} - \beta_\mathrm{S})}{\beta_\mathrm{L}}\frac{2P_\mathrm{S}\frac{RH+R^2}{r^2}}{P_\mathrm{L}\frac{\pi R^2 H}{\frac{4}{3}\pi r^3} - P_\mathrm{S}\frac{RH+R^2}{r^2}}$$

The second term in the denominator can be neglected for all practical purpose

$$\begin{aligned}\frac{T_\mathrm{mp,nano}}{T_\mathrm{mp,bulk}} &= 1 - \frac{(\beta_\mathrm{L} - \beta_\mathrm{S})}{\beta_\mathrm{L}}\frac{8P_\mathrm{S}\frac{RH+R^2}{r^2}}{3P_\mathrm{L}\frac{R^2H}{r^3}} = 1 - \frac{(\beta_\mathrm{L} - \beta_\mathrm{S})}{\beta_\mathrm{L}}\frac{8P_\mathrm{S}\frac{HR^2}{r^2}\left(\frac{1}{H}+\frac{1}{R}\right)}{3P_\mathrm{L}\frac{R^2H}{r^3}} \\ &= 1 - \frac{(\beta_\mathrm{L} - \beta_\mathrm{S})}{\beta_\mathrm{L}}\frac{8}{3}\frac{P_\mathrm{S}}{P_\mathrm{L}}r\left(\frac{1}{H}+\frac{1}{R}\right)\end{aligned} \tag{13}$$

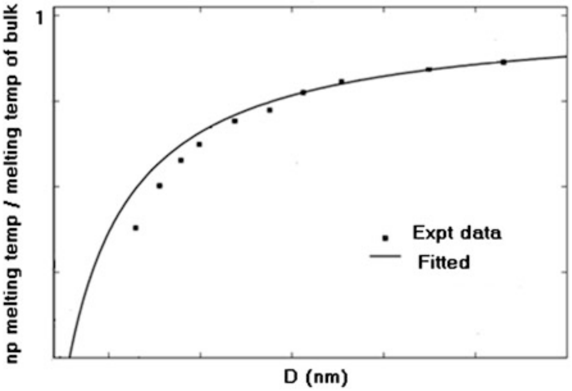
In the case of nanowire, it may be assumed that $H \gg R$, i.e., $\frac{1}{R} \gg \frac{1}{H}$, then the melting point may be written as

$$\frac{T_\mathrm{mp,nano}}{T_\mathrm{mp,bulk}} = 1 - \frac{(\beta_\mathrm{L} - \beta_\mathrm{S})}{\beta_\mathrm{L}}\frac{8}{3}\frac{P_\mathrm{S}}{P_\mathrm{L}}\frac{r}{R} \tag{14}$$

So we may conclude that the melting point of nanowire decreases inversely to its radius (R). Experimentally measured melting points for nanosheets, fitted with Eq. (14) for different value of "q," are shown in Fig. 1.

For nanosheet (also called nanofilm), it may be approximated that $R \gg H$ i.e., $\frac{1}{H} \gg \frac{1}{R}$. So for such nanosheets Eq. (13) may be written as

Fig. 1 Dependence of melting temperature on the diameter ($D = 2R$) of isolated In nanoparticles

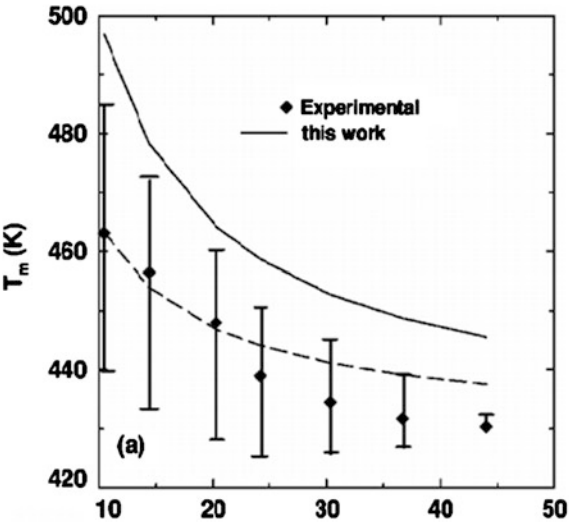


$$\frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} = 1 - \frac{(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})}{\beta_{\mathrm{L}}} \frac{8}{3} \frac{P_{\mathrm{S}}}{P_{\mathrm{L}}} \frac{r}{H} \tag{15}$$

So, for these systems, we may conclude that melting point deceases inversely to their height (H). Experimentally measured melting points, fitted with Eq. (15) for different value of "q," are shown in Fig. 2.

Though the variation of melting point of nanoparticles, nanowires, nanosheets can be well explained on the basis of Eq. (3), but it has some limitations. For example, this present theory cannot be used to explain melting point of system containing more than one element or nanoparticles embedded in matrix. If we consider second case, then the number of bonds that the surface atoms makes with matrix elements is $n_{\mathrm{s}}(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})$ and the corresponding energy of the nanoparticle may be written as

Fig. 2 Superheating of In nanoparticles embedded within Al matrix. *Dots* represent experimental points. *Line* shows the theoretically fitted curve (Reprinted with permission from Nanda et al. [2] © American Physical Society 2002)



$$E_{\mathrm{c,nano,matrix}} = \frac{1}{2}n_{\mathrm{i}}\beta_{\mathrm{L}}\varepsilon + \frac{1}{2}n_{\mathrm{s}}\beta_{\mathrm{S}}\varepsilon + n_{\mathrm{s}}(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})\varepsilon_{\mathrm{m}} \tag{16}$$

where ε_{m} is the bond energy between surface and matrix atoms. Thus the melting point for such nanoparticles can be written according to Eq. (3),

$$\begin{aligned}\frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} = \frac{E_{\mathrm{c,nano,matrix}}}{E_{\mathrm{c,bulk}}} &= \frac{\frac{1}{2}n_{\mathrm{i}}\beta_{\mathrm{L}}\varepsilon + \frac{1}{2}n_{\mathrm{s}}\beta_{\mathrm{S}}\varepsilon + \frac{1}{2}n_{\mathrm{s}}(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})\varepsilon_{\mathrm{m}}}{\frac{1}{2}n_{\mathrm{t}}\beta_{\mathrm{L}}\varepsilon} \\ &= 1 - \frac{(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})}{\beta_{\mathrm{L}}}\frac{n_{\mathrm{s}}}{n_{\mathrm{t}}} + 2\frac{(\beta_{\mathrm{L}} - \beta_{\mathrm{S}})}{\beta_{\mathrm{L}}}\frac{n_{\mathrm{s}}}{n_{\mathrm{t}}}\frac{\varepsilon_{\mathrm{m}}}{\varepsilon}\end{aligned} \tag{17}$$

If $\varepsilon_{\mathrm{m}} \gg \varepsilon$, then there may be certain situation such that $2\frac{(\beta_{\mathrm{L}}-\beta_{\mathrm{S}})}{\beta_{\mathrm{L}}}\frac{n_{\mathrm{s}}}{n_{\mathrm{t}}}\frac{\varepsilon_{\mathrm{m}}}{\varepsilon} > \frac{(\beta_{\mathrm{L}}-\beta_{\mathrm{S}})}{\beta_{\mathrm{L}}}\frac{n_{\mathrm{s}}}{n_{\mathrm{t}}}$. In this situation, $T_{\mathrm{mp,nano}}$ is greater than $T_{\mathrm{mp,bulk}}$, i.e., melting point increases in comparison to melting point of bulk system. This phenomenon of enhancement of melting point is called superheating of the nanoparticles.

There also exist different methods to illustrate the melting point of nanoparticles. One of them is liquid-drop model [2]. According to this model, $E_{\mathrm{c,nano}}$ of a nanoparticle consisting of N number of atoms can be written as

$$E_{\mathrm{c,nano}} = a_v N - 4\pi r^2 N^{\frac{2}{3}}\gamma \tag{18}$$

where a_v and γ represent the cohesive energy of an interior atom and surface tension of the materials in isolated condition, respectively. First term represents the cohesive energy of particle neglecting surface effects and second term represent surface effects on the cohesive energy of the nanoparticle. So the melting point of the nanoparticle may be written as according to Eq. (3),

$$\frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} = \frac{a_v N - 4\pi r^2 N^{\frac{2}{3}}\gamma}{a_v N} = 1 - \frac{4\pi r^2 N^{-\frac{1}{3}}\gamma}{a_v} \tag{19}$$

Considering volume of the nanoparticle is sum of the volume of individual atoms, and then $N^{-\frac{1}{3}} = \frac{r}{R}$. Substituting for $N^{-\frac{1}{3}}$ into Eq. (19), we may write the following equation

$$\frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} = 1 - \frac{4\pi r^3\gamma}{a_v}\frac{1}{R} \tag{20}$$

When the nanoparticles are embedded within some matrix, then surface atoms of the nanoparticles make some bonds with the matrix elements and then the surface tension is modified to $\gamma - \alpha\gamma_{\mathrm{m}}, \gamma_{\mathrm{m}}$ represent the coefficient surface energy of the surrounding material and α is the correlation between atoms of the nanoparticles and those of the surrounding matrix; $\alpha = 0$ for isolated free nanoparticles and $\alpha = 1$

for epitaxy between nanoparticles and the surrounding matrix. Using this modified expression for surface energy we obtain

$$\frac{T_{\mathrm{mp,nano}}}{T_{\mathrm{mp,bulk}}} = 1 - \frac{4\pi r^3}{a_v} \frac{\gamma - \alpha\gamma_{\mathrm{m}}}{R} \tag{21}$$

So from above equation, it may be noted that if $\gamma < \alpha\gamma_{\mathrm{m}}$, then the nanoparticles will melt above the bulk melting point, i.e., melting point of nanoparticle, embedded in matrix will increase. This liquid-drop model is in well agreement with the experimental result of superheating of In and Pb nanoparticles embedded in Al matrix [3–6].

3 Electronic Structure and Optical Property

Optical properties of semiconductor nanostructures have gained special interest because they can bridge the gap between bulk and atom and very often possesses new avenue of application especially in optoelectronics, solar cell, electronics, light emitting diodes, etc. The optical properties of these nanostructures are determined by their electronic structure that significantly differs for various morphologies since electronic structure of the nanomaterials are very much dependent on surface atoms. Size effects on optical properties are observed within particles having size ~ 10 nm. In recent time, quantum structures have also particularly drawn special attention because some of other fundamental properties like dielectric properties and electrical conductivity are found to be tuneable by their size. Interestingly, quantum structures are characterized by staircase-like *I–V* characteristic curve instead of linear *I–V*, atom-like optical line spectra, etc. They also possess exciting properties in optical absorption, electron-hole recombination, luminescence, etc., completely different from bulk. Such exciting properties of these nanostructure materials arise mainly from quantum confinement effect that modifies the band structure and the corresponding DOS of the materials. Commonly, quantum effect is found to be prominent when energy-level spacing between two successive states becomes greater than the thermal energy. As an example, it can be said that in bulk semiconducting materials, energy levels are continuous, but in the case of nanostructures, energy levels become discrete, i.e., institutively we may conclude that we have delocalized wave function of the electrons in one side (in bulk), but on the other side wave function of the electrons has discrete nature mostly atomic like (in quantum dot). Thus main advantage is that by changing the size of the nanoparticle, we may tune wave function of the electrons to a great extent and the corresponding properties of the materials vary over a wide range. Not only tuning, the nanostructure materials also exhibits several new properties depending on the surface atoms.

Before going into details, let us review some general features of the electronic structure of material. In the early stage of materials science, classical kinetic theory was adopted to illustrate all the physics properties like of materials. But that motivation was in vain, because the theoretically predicted properties and their variation were not supported by experimental results. For example, according to classical theory, electronic contribution to specific heat in the case of metal was found to be independent of temperature in the low temperature region, but experimentally is was found to be proportional to absolute temperature (T). After development of quantum mechanics, electrons were assumed be free particle within the materials, i.e., electron–lattice and electron–electron interaction interactions were neglected (Sommerfeld theory of free electron theory). Then Schrödinger equation for free electrons was solved and the subsequently developed consequences were used to explain the linear temperature dependence of specific heat. Similar to three quantum numbers (excluding spin quantum number) that is used to specify states of an electron in the case of atom, three quantum numbers, associated three components of a wave vector corresponding to three directions in space, are used to specify electronic states. But such free electron theory was not able to introduce band gap as in the case of semiconducting materials. Then electron–lattice interaction was taken into account to solve Schrödinger equation (Bloch theory). Since lattice is periodically arranged, so the electron–lattice interaction, presented by potential function in the Schrödinger equation, is periodic in nature. In the presence of periodic potential, wave function of the electrons is also found to be periodic. These electrons are called Bloch electron after the name of famous Swiss physicist Felix Bloch. Solution of Schrödinger equation for Bloch electrons was carried out by Krönig–Penny. They showed that the allowed k-values are limited in some region, i.e., allowed energy levels are also limited. Energy levels form band within their allowed levels. Up to some energy bands, states are occupied by electron, above which they are empty. In this context, energy bad gap (E_g) is defined as the energy difference between highest occupied energy state, known as valence band and lowest unoccupied energy bands, called conduction band. Therefore, the introduction of periodic potential opens the concept of energy band gap and the main idea is to calculate energy dispersion relation, i.e., the relation between energy states and wavevectors ($E(k)$ diagram). For details of the calculation procedure, one may consult fundamental books on solid-state physics or materials science. Later, some advanced techniques like linear combination of atomic orbitals, augmented plane wave method (APW) were developed to calculate more realistic electronic structure of materials. Very recently, a completely new technique, known as density functional theory, was developed by Kohn–Sham for calculation of band structure of any materials. This technique relies on electron density to calculate ground state energy and their wave function. Band structure of one of the simplest semiconducting materials Si, is presented in Fig. 3.

If we look institutively at the highest energy point in the valence band or lowest energy point in conduction band, dispersive nature is mostly parabolic, i.e., we may write energy dispersive relation $E = \frac{\hbar^2}{2m_{\text{eff}}}k^2$ for electronic states. This is the

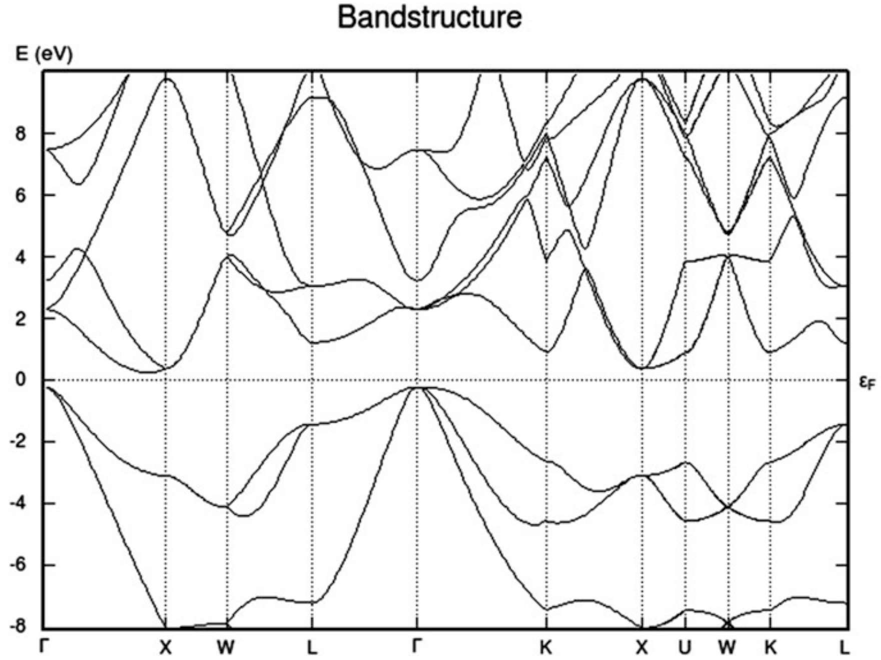


Fig. 3 Band structure of Si

dispersion relation same as free electron, thus their behavior is supposed to be like free electron. One point that has to be mentioned here is that the mass of the electron that we are considering here is effective mass (m_{eff}) of an electron instead of free electronic mass. m_{eff}, given by the equation $m_{\text{eff}} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2}$, depends on the curvature of the bands, hence it varies from material to material. Since electrons are free, we may approximate $V = 0$ for these electrons and they are confined within the materials, i.e., they experience practically infinite potential at the surface. Thus we may treat the problem as free particle in three-dimensional infinite square well potential. To explain different physical property related to electronic structure, the important parameter that is defined is the DOS, $D(E)$. The significance of the $D(E)$ is that it represents the number of states having energy between E and $E + \mathrm{d}E$.

For bulk system, DOS is calculated to be

$$\mathrm{D}(E) = \frac{1}{2\pi^2}\left(\frac{2m_{\text{eff}}}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}} \tag{22}$$

Details of the calculation can be found in any books in solid-state physics. For materials having size in the nanodimension, different physical properties those are deviating from bulk mainly originate from size-modified dispersive relation and $D(E)$. Commonly nanostructures are classified into (a) two-dimensional, e.g., quantum well, (b) one-dimensional, e.g., nanowire and (c) zero-dimensional, e.g., quantum dots. So to understand the electronic structure and associated optical properties of different nanostructures, primarily it is mandatory to understand their $D(E)$.

Quantum well structure is defined as the structures that restrict motion of electron in one direction (let z-direction) whereas in other two directions (x, y) it is free. So it is free movement in the x, y plane is entirely characterized k_x and k_y. To calculate DOS of such quantum structures, we consider a circular area in k-space with radius $k = \sqrt{k_x^2 + k_y^2}$. Therefore, associated area of the circle $A_k = \pi k^2$ and area of each state (A_{state}) is $A_{\text{state}} = k_x \cdot k_y$. With $k_x = \frac{2\pi}{L_x}$ and $k_y = \frac{2\pi}{L_y}$, where L_x and L_y are length of the structure along x and y directions, respectively, we may obtain $A_{\text{state}} = \frac{4\pi^2}{L_x \cdot L_y}$. Thus total number of states (N), encompassed by the circle, can be written as $N = \frac{2A_k}{A_{\text{state}}}$, here factor 2 has been introduced for spin degeneracy. Considering dispersion relation for free electrons and $L_x \cdot L_y = 1$, we may write the following expression for N,

$$N = \frac{m_{\text{eff}} E}{\pi \hbar^2} \tag{23}$$

Thus $D(E)$ for such quantum well structure, i.e., number of states having energy between E and E + dE,

$$D(E) = \frac{m_{\text{eff}}}{\pi \hbar^2} \tag{24}$$

The important point that must be noticed from the above expression for $D(E)$ is that it found to be independent of energy. Other important point that has to be discussed is that the energy (E), we are considering here, is not the total energy (E_{total}) of the electron. Therefore, is another component of associated with z-direction of motion of the electron; let it be E_z. Next our unresolved problem is to find out the expression for E_z. Considering electron near band edge as free, Schrödinger equation for movements of electrons in this direction can be written as

$$\frac{-\hbar^2}{2m_{\text{eff}}} \frac{\partial^2 \varphi(z)}{\partial z^2} = E\varphi(z) \tag{25}$$

With boundary condition $V = \infty$ at $Z = 0$ and L_z, L_z represents the height of the quantum well along z-direction. The problem is very much similar with the particle in one-dimensional infinite square well potential. So the solution of $\varphi(z)$, i.e., normalized wave function of the electron in this direction may be written as,

$$\varphi(z) = \frac{\sqrt{2}}{L_z} \sin\left(\frac{n_z \pi}{L_z} z\right), \quad \text{where } n_z = 1, 2, 3\ldots \tag{26}$$

Their energy can be written as

$$E_z = \frac{\hbar^2}{2m_{\text{eff}}} \left(\frac{n_z \pi}{L_z}\right)^2 \tag{27}$$

E_z is characterized by the quantum number wave vector n_z. Therefore, we will use suffix n_z to represent E_z, let it presented by E_{n_z}. So, in this situation DOS can be written as

$$D(E) = \frac{m_{\text{eff}}}{\pi \hbar^2} \sum_{n_z} \Theta(E_{\text{total}} - E_{n_z}) \tag{28}$$

where $\Theta(E_{\text{total}} - E_{n_z})$ is the Heaviside unit step function, defined by

$$\Theta\left(E_{\text{total}} - E_{n_z}\right) = \begin{cases} 0 & \text{if } E_{\text{total}} < E_{n_z} \\ 1 & \text{if } E_{\text{total}} > E_{n_z} \end{cases} \tag{29}$$

Therefore, from above equations it is clear that E_{n_z} and subsequently D (E) depend on L_z. Recent development of the synthesis techniques for nanomaterials enables us to have better control over synthesis of nanomaterials in terms of their size and shape. For example, quantum well of any desired size can be fabricated by manipulating atoms on atomically flat surface using scanning tunneling microscope. Thus we have technology that can be used to tailor wave function of the electrons and their corresponding energy levels. All the properties of the material mainly depend on the electronic wave function and their associated energy levels, thus it becames possible to get desired property by fabricating artificial nanostructures with suitable well structure.

Nanowire confines electrons in two dimensions (let they be y and z), whereas they are free along direction of length of the wire (let it be x). Then delocalized state of the electron, in the x direction, is characterized by wave vector $k_x = \frac{2\pi}{L_x}$, where L_x represents length of the nanowire. Similar to previous quantum well structures, total of length of k-space within this structure is $L_k = 2k$ and length of each k-point $L_{\text{state}} = \frac{2\pi}{L_x}$. Thus total number of k-points (N) considering spin degeneracy can be written as $N = \frac{2L_k}{L_{\text{state}}}$. From dispersive relation corresponding to free electron, we may write the following expression for N (L_x = 1),

$$N = \frac{2}{\pi} \frac{\sqrt{2m_{\text{eff}} E}}{\hbar} \tag{30}$$

Therefore $D(E)$ for such nanowire can be written as

$$D(E) = \frac{1}{\pi}\frac{\sqrt{2m_{\text{eff}}}}{\hbar}\frac{1}{\sqrt{E}} \tag{31}$$

Following similar arguments corresponding to well structure, E is the difference between total energy and confining energy in other two direction ($E_{y,z}$), i.e., $E = E_{\text{total}} - E_{y,z}$. Similar to well structure, we may obtain, after solving Schrödinger equation, the quantized value of $E_{y,z}$ and they are characterized by two quantum number (n_y and n_z). So, we will use suffix n_y and n_z to denote them (let it be denoted by E_{n_y,n_z}).

Therefore, $D(E)$ in terms of total energy may be written for nanowire structure as

$$D(E) = \frac{1}{\pi}\frac{\sqrt{2m_{\text{eff}}}}{\hbar}\sum\nolimits_{n_y,n_z}\frac{1}{\sqrt{E_{\text{total}} - E_{n_y,n_z}}}\Theta(E_{\text{total}} - E_{n_y,n_z}) \tag{32}$$

The important point to be noted here is that $D(E)$ is inversely proportional square root of energy.

In a similar fashion, we may find the general expression of $D(E)$ corresponding to a perfect quantum dot,

$$D(E) = 2\sum_{n_x,n_y,n_z}\delta(E - E_{n_x,n_y,n_z}) \tag{33}$$

Not only the modification of DOS and quantization of energy states are observed for nanoparticles, they also exhibit completely different type of states originating from surface atoms. In the case of bulk materials, since the number of surface atom is negligible compare to interior atom, their properties are entirely determined by the interior atoms. But in the case of nanoparticles, surface atom contribution increases as particle size gets reduced, hence properties of nanoparticles are significantly affected by surface atoms. In many cases, surface atoms possess exotic localized electronic states and very often play a crucial role in determining properties of the material mainly optical properties. Calculation shows that the wave function corresponding to these localized states is exponentially decaying in nature (since, in this region potential energy is greater than total energy of the particle) and energy associated with these states lies in the band gap. These kinds of localized states are known as “Shockley states.” In addition to this, it must be mentioned that localization of electronic is also possible for any local modulation of crystalline potential. The electronic states with this origin are known as “Tamm states.” Most of the optical properties of extremely small particles are determined by these surface localized states. In order to obtain some more intuitive idea about the surface localized states, we consider a bulk crystalline material of group IV such as C, Si, Ge, etc., having diamond structure. Their stable structure is constructed via the formation of interatomic bonds from hybridization between sp^3 orbitals of neighboring atoms. After hybridization, they give either bonding state or antibonding state. The bonding states are fully occupied whereas antibonding states are empty and energy difference between them is defined as band gap of the corresponding material. Energy of the bonding state is found to lower than that of unhybridized state, whereas antibonding state has higher energy. The formation of interatomic bonds and splitting of sp^3 orbitals occur everywhere within the material, i.e., bonding and antibonding states are highly degenerate. However for an atom at the surface of the nanoparticles, it has no neighbor at the vacuum side. As a result, the unpaired bond of this surface atom with intrinsic nature of sp^3 hybridized state intrudes toward the vacuum side and energy corresponding to this lies between bonding and antibonding states. These unpaired bonds are called dangling bond. These dangling bonds play a crucial role in the properties of the nanoparticles. These unfavorable surface (since their energy is greater than bonding state) states also act as a driving force to determine the shape of the nanoparticles. For example, the nanoparticle with isotropic surface energy always grows with spherical shape. This happens because sphere is the shape with smallest surface area under a volume. On the other hand, for anisotropic surface energy, nanoparticles grow with polyhedron shape with specific surfaces corresponding to low surface energy (Fig. 4).

Absorption and emission of light by semiconducting nanoparticles are the two important properties those are generally used to study their size-dependent optical property. Before going into details of the size effect on the optical properties, let us have quick look on light–matter interaction on the basis of optical transition.

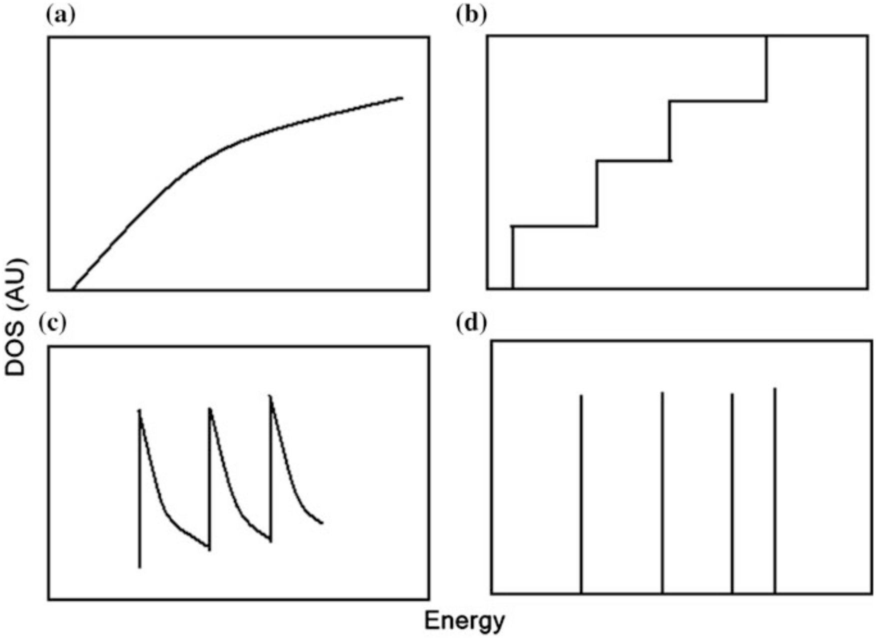


Fig. 4 Density of state; **a** bulk system, **b** two-dimensional quantum well, **c** one-dimensional nanowire, **d** zero-dimensional quantum dot

Classical theory on light–matter interaction started with Lorentz oscillator model that correctly describes the interaction on the basis of real and imaginary part of the refractive index. After development of quantum mechanics, a semiclassical theory was developed to quantify the interaction. In this approach, atoms or molecules, involved in this process, are treated quantum mechanically considering time-dependent Schrödinger equation and radiation was treated classically on the basis of Maxwell's equation. This method well explains the absorption and emission of light. Here, light emission or absorption process involves two discrete quantum states of an electron (with energy E_1 and $E_{2,}$ respectively, $E_2 > E_1$) corresponding to atoms or molecules. Let $\Psi_1(\mathbf{r},t) = \psi_1(\mathbf{r})\exp\left(-\frac{iE_1}{\hbar}t\right)$ and $\Psi_2(\mathbf{r},t) = \psi_2(\mathbf{r})\exp\left(-\frac{iE_2}{\hbar}t\right)$ be the two time-dependent electronic waves corresponding to these two states. $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ are the time-independent part of the wave function. If "H" be the Hamiltonian of the system describing the system and the wave functions, we are using here, be the eigenfunction of H, then $H\Psi_1(\mathbf{r},t) = E_1\Psi_1(\mathbf{r},t)$ and $H\Psi_2(\mathbf{r},t) = E_2\Psi_2(\mathbf{r},t)$. Since the two wave functions are the eigenfunction of the same Hamiltonian, they are orthogonal to each other. The general wave function ($\Psi(\mathbf{r},t)$), representing the system is the superposition of these two wave functions, i.e., $\Psi(\mathbf{r},t) = C_1\Psi_1(\mathbf{r},t) + C_2\Psi_2(\mathbf{r},t)$. Since the wave function is time dependent, so the constants C_1 and C_2 will also be time dependent. If $\Psi(\mathbf{r},t)$ is considered to be normalized wave function, the constants $C_1(t)$ and $C_2(t)$ should satisfy the relation $|C_1(t)|^2 + |C_2(t)|^2 = 1$. In the presence of electromagnetic radiation with electric field $\mathbf{E} = \mathbf{E}_0\exp^{i(\mathbf{k}\cdot\mathbf{r}+\omega t)}$, the additional potential energy (V_{rad}) that acts on the electron is $V_{\text{rad}} = -e\mathbf{E}_0\exp^{i(k\cdot r-\omega t)} = V_{\text{rad}}(\mathbf{r})\exp^{i\omega t}$. So the Hamiltonian of the system in the presence of radiation may be written as $H_{\text{total}} = H + V_{\text{rad}}$. If we consider electron with any given wave function (for particular choice of $C_1(t)$ and $C_2(t)$), in the presence of V_{rad} it will be modified. Since $\Psi_1(\mathbf{r},t)$ and $\Psi_2(\mathbf{r},t)$ are two basis wave functions, so the modified wave function may also be written as the linear combination of these basis wave functions (with different combination of C_1 and C_2), but it will satisfy the modified Schrödinger equation, i.e.,

$$i\hbar\frac{\partial}{\partial t}(C_1(t)\Psi_1(\mathbf{r},t) + C_2(t)\Psi_2(\mathbf{r},t)) = H_{\text{total}}(C_1(t)\Psi_1(\mathbf{r},t) + C_2(t)\Psi_2(\mathbf{r},t)) \quad (34)$$

Using $H\Psi_1(\mathbf{r},t) = E_1\Psi_1(\mathbf{r},t)$ and $H\Psi_2(\mathbf{r},t) = E_2\Psi_2(\mathbf{r},t)$, the above equation may be simplified to

$$i\hbar\left(\Psi_1(\mathbf{r},t)\frac{\mathrm{d}C_1(t)}{\mathrm{d}t} + \Psi_2(\mathbf{r},t)\frac{\mathrm{d}C_2(t)}{\mathrm{d}t}\right) = V_{\text{rad}}(C_1(t)\Psi_1(\mathbf{r},t) + C_2(t)\Psi_2(\mathbf{r},t)) \quad (35)$$

If we multiply the above equation by $\Psi_1(\mathrm{r},t)^*$ and do the integration all over that space, then from orthogonality condition ($\int\Psi_1(\mathbf{r},t)^*\Psi_2(\mathbf{r},t)\mathrm{d}r = 0$),

$$i\hbar\frac{\mathrm{d}C_1(t)}{\mathrm{d}t} = C_1(t)\int \Psi_1(\mathbf{r},t)^* V_{\mathrm{rad}}\Psi_1(\mathbf{r},t)\mathrm{d}\boldsymbol{r} + C_2(t)\int \Psi_1(\mathbf{r},t)^* V_{\mathrm{rad}}\Psi_2(\mathbf{r},t)\mathrm{d}\boldsymbol{r} \tag{36}$$

or,

$$i\hbar\frac{\mathrm{d}C_1(t)}{\mathrm{d}t} = C_1(t)\int \psi_1(\boldsymbol{r})^* V_{\mathrm{rad}}\psi_1(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + C_2(t)\exp\left(-i\left(\frac{(E_2-E_1)}{\hbar} - \omega\right)t\right)\int \psi_1(\boldsymbol{r})^* V_{\mathrm{rad}}(r)\psi_2(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{37}$$

Using the expression of $V_{\mathrm{rad}} = V_{\mathrm{rad}}(\mathbf{r})\exp^{i\omega t}$ in the above equation,

$$i\hbar\frac{\mathrm{d}C_1(t)}{\mathrm{d}t} = C_1(t)\int \psi_1(\boldsymbol{r})^* V_{\mathrm{rad}}(\mathbf{r})_{\mathrm{rad}}\psi_1(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + C_2(t)\exp\left(-i\left(\frac{(E_2-E_1)}{\hbar} - \omega\right)t\right)$$
$$\int \psi_1(\boldsymbol{r})^* V_{\mathrm{rad}}(\mathbf{r})\psi_2(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{38}$$

Similarly, we may write the time dependence equation for $C_2(t)$,

$$i\hbar\frac{\mathrm{d}C_2(t)}{\mathrm{d}t} = C_2(t)\int \psi_2(\boldsymbol{r})^* V_{\mathrm{rad}}(\mathbf{r})_{\mathrm{rad}}\psi_2(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + C_1(t)\exp\left(-i\left(\frac{(E_2-E_1)}{\hbar} - \omega\right)t\right)$$
$$\int \psi_2(\boldsymbol{r})^* V_{\mathrm{rad}}(\mathbf{r})\psi_1(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{38}$$

The equation may also be written as

$$i\hbar\frac{\mathrm{d}C_2(t)}{\mathrm{d}t} = C_2(t)\int \psi_2(\boldsymbol{r})^* V_{\mathrm{rad}}(\mathbf{r})_{\mathrm{rad}}\psi_2(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + C_1(t)\exp\left(-i\left(\frac{(E_2-E_1)}{\hbar} - \omega\right)t\right)M_{12} \tag{39}$$

With $M_{12} = \int \psi_2(\boldsymbol{r})^* V_{\mathrm{rad}}(\mathbf{r})\psi_1(\boldsymbol{r})\mathrm{d}\boldsymbol{r}$, If we look at the Eq. (39) intuitively, the absolute square of the coefficients gives us the time dependence of the occupation probability of electrons in these states. To make the problem more simple, we assume that the initially the electron is completely described by the wave function $\psi_1(\boldsymbol{r})$, i.e., $C_1(t=0)=1$ and $C_2(t=0)=0$. At later time, in the presence of radiation, there will be nonzero occupation probability at other state. In this situation, $C_1(t\neq 0)\sim 1$, but $C_2(t\neq 0)\neq 0$, i.e., electrons are being excited to and excited from ground state (E_1) to higher energy state (E_2). At this time variation of $C_1(t\neq 0)$ is not so significant, but the variation of $C_2(t\neq 0)$ is remarkable. So the transition rate is given by time dependence of $C_2(t)$ that simplified into the following equation,

$$C_2(t) = \frac{-i}{\hbar} M_{12} \int \exp\left(-i\left(\frac{(E_2 - E_1)}{\hbar} - \omega\right)t\right) \mathrm{d}t \tag{40}$$

If we do the integration over time period (to obtain average value), it will be nonzero only when $E_2 - E_1 - \hbar\omega = 0$, i.e., $E_2 = E_1 + \hbar\omega$. Detail calculation may be found in any book of basic quantum mechanics. Substituting the value of $V_{\mathrm{rad}}(\mathbf{r}) = -e\mathbf{E}_0 \exp^{i\mathbf{k}\cdot\mathbf{r}}$ into the expression of M_{12}, it possible to evaluate M_{12},

$$M_{12} = -e\mathbf{E}_0 \int \psi_2(\boldsymbol{r})^* \exp^{i\mathbf{k}\cdot\mathbf{r}}(\mathbf{r}) \psi_1(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \tag{41}$$

$|M_{12}|$ is called transition dipole moments of the electronic transition between $\psi_1(\boldsymbol{r})$ and $\psi_2(\boldsymbol{r})$. The significance of Eq. (41) is that it gives the transition probability between two states, provided their energy difference is equal to energy of incident electromagnetic radiation. If $\psi_1(\boldsymbol{r})$ is supposed to be the wave function on an electron in valence band maxima and $\psi_2(\boldsymbol{r})$ be that of an electron in conduction band minima, then $|M_{12}|$ illustrates the transition probability of an electron from valence band maxima to conduction band minima and their energy difference provided their energy difference, i.e., band gap is equal to the energy corresponding to incident radiation. These two are the fundamental relation describing optical absorption and emission process. Wave functions those we are considering here must be different for different materials, i.e., transition matrix depends on the materials. Hence, absorption coefficient of the materials that proportional to transition matrix becomes material dependent. Once we excite an electron to conduction band, after certain characteristic time, known as radiative life time, it will again come to valence band after emitting energy equal to band gap energy. If band gap energy lies within our visible region, the emission process will be associated with some visible color. Depending on the life time, the emission process is classified into two categories: fluorescence and phosphorescence.

Experimentally, it is proved that band gap of semiconductor nanoparticles increases if its size is reduced, i.e., abruption edge shifts toward lower wavelength. Similarly, during emission, the peak in the emission spectra also shifts toward lower wavelength. This shift is known as Blue shift. If emission spectrum lies within the visible region, then the corresponding color of the material may change upon size reduction (shown in Fig. 5). To understand these experimental results theoretically, we consider electronic states near valence band maxima and conduction band minima those are mainly responsible for absorption and emission process. Near edge states, electrons are considered as free electrons, i.e., energy dispersive relation for them is given by $E = \frac{\hbar^2}{2m_{\mathrm{eff}}}k^2$. So these electronic states are characterized by wave vector k. We consider an electronic state just below the valence band maxima (k is very low, but no equal to zero), and then its energy ($E_{k,\mathrm{VB}}$) is given by

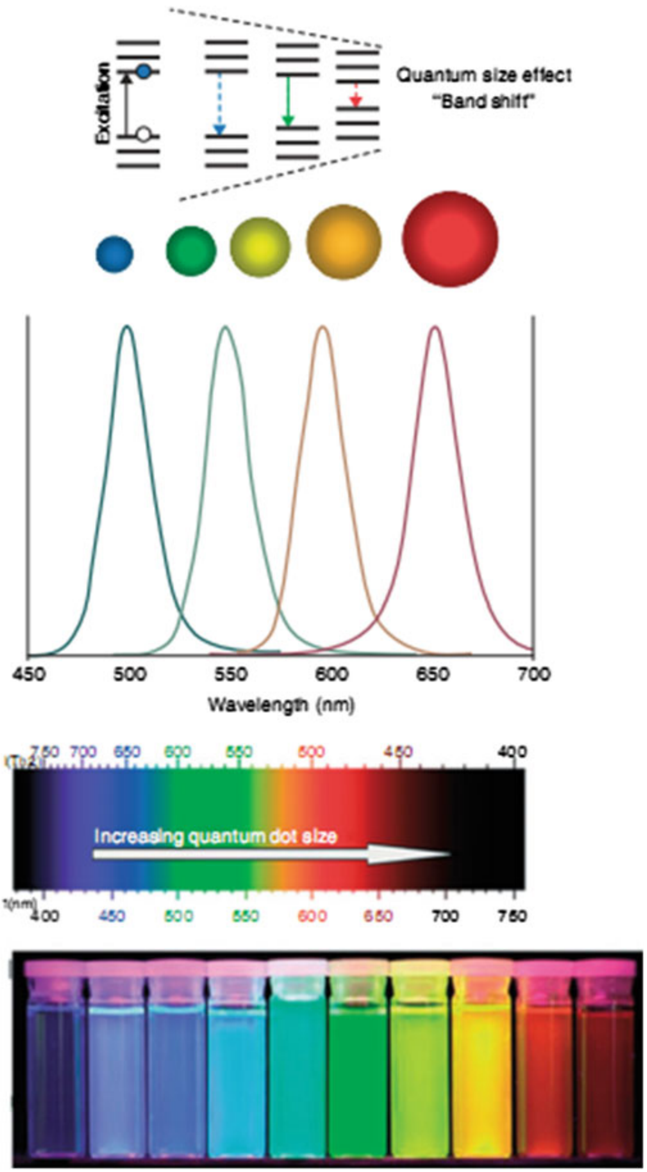


Fig. 5 Size-dependent PL colors of semiconductor QDs. Electronic structure of QDs correlated with the QD radius and resulting 'blue shift' due to quantum confinement as the QD size decreases. Reproduced with permission from Ref. Mansur [19]. Copyright: Wiley VCH

$$E_{k,\mathrm{VB}} = E_{\mathrm{VBM}} - \frac{\hbar^2}{2m_{\mathrm{h,eff}}}k^2 \tag{42}$$

Similarly, the energy of the electronic state just above conduction band minima ($E_{k,\mathrm{CB}}$) can be written as

$$E_{k,\mathrm{CB}} = E_{\mathrm{CBM}} + \frac{\hbar^2}{2m_{\mathrm{e,eff}}}k^2 \tag{43}$$

where E_{VBM} and E_{CBM} represents the energy associated with valence band maxima and conduction minima, respectively. The energy difference between them (E_k),

$$E_k = E_{k,\mathrm{CB}} - E_{k,\mathrm{VB}} \tag{44}$$

Or

$$E_k = E_{\mathrm{CBM}} + \frac{\hbar^2}{2m_{\mathrm{e,eff}}}k^2 - E_{\mathrm{VBM}} + \frac{\hbar^2}{2m_{\mathrm{h,eff}}}k^2 \tag{45}$$

$$E_k = E_{\mathrm{CBM}} - E_{\mathrm{VBM}} + \frac{\hbar^2}{2}\left(\frac{1}{m_{h,\mathrm{eff}}} + \frac{1}{m_{\mathrm{e,eff}}}\right)k^2 \tag{46}$$

Now in the case of semiconducting nanoparticles, electrons are free and are spatially confined within a narrow region space. So we may use "particle in box" model in three dimensions for them. From quantization of k that comes from confinement condition, we may write $k = \frac{\pi}{R}$. Using this condition into above equation, we may write

$$E_k = E_{\mathrm{CBM}} - E_{\mathrm{VBM}} + \frac{\hbar^2}{2}\left(\frac{1}{m_{\mathrm{h,eff}}} + \frac{1}{m_{\mathrm{e,eff}}}\right)\left(\frac{\pi}{R}\right)^2 \tag{47}$$

Thus from above equation, it is clear that the energy difference between two states with same k value increases as we decrease the particle size. Extending this concept to band gap of semiconducting nanoparticles, we may conclude that it will increase if size of the nanoparticle is reduced. So in the case of band gap of semiconducting materials, we may write for band gap of nanoparticles ($E_{\mathrm{g,nano}}$),

$$E_{\mathrm{g,nano}} = E_{\mathrm{g,bulk}} + \frac{\hbar^2}{2}\left(\frac{1}{m_{\mathrm{h,eff}}} + \frac{1}{m_{\mathrm{e,eff}}}\right)\left(\frac{\pi}{R}\right)^2 \tag{48}$$

The above equation is one of the fundamental for nanoparticles illustrating that the band gap will increase if the size of the particle gets reduced. In emission spectra, it is very often observed that in addition to band to band transition, new emission spectrum arises at higher wavelengths in the case of extremely small particle size. The effect is prominent in ZnO quantum dots. This emission is mainly attributed to the surface related Shockley defects states.

Once electron is being excited to conduction band, it leaves a hole in the valence band. Before recombination, the electron spends some time that is the characteristic life time of the electron in conduction band for that particular material, known as radiative lifetime; the excited electron interacts with hole. The interaction between them is always attractive since they carry opposite charge (Coulomb force). Depending on the dielectric constant of the material, they may either form a bound state (in case of materials with low dielectric constant) or may not (in case of materials with high dielectric constant). If they do not form bound state, then electron in conduction band and hole in valence band move independently. On the other hand, if they form bound state, the electronic excitation is described by a bound electron–hole pair having energy inversely proportional to the dielectric

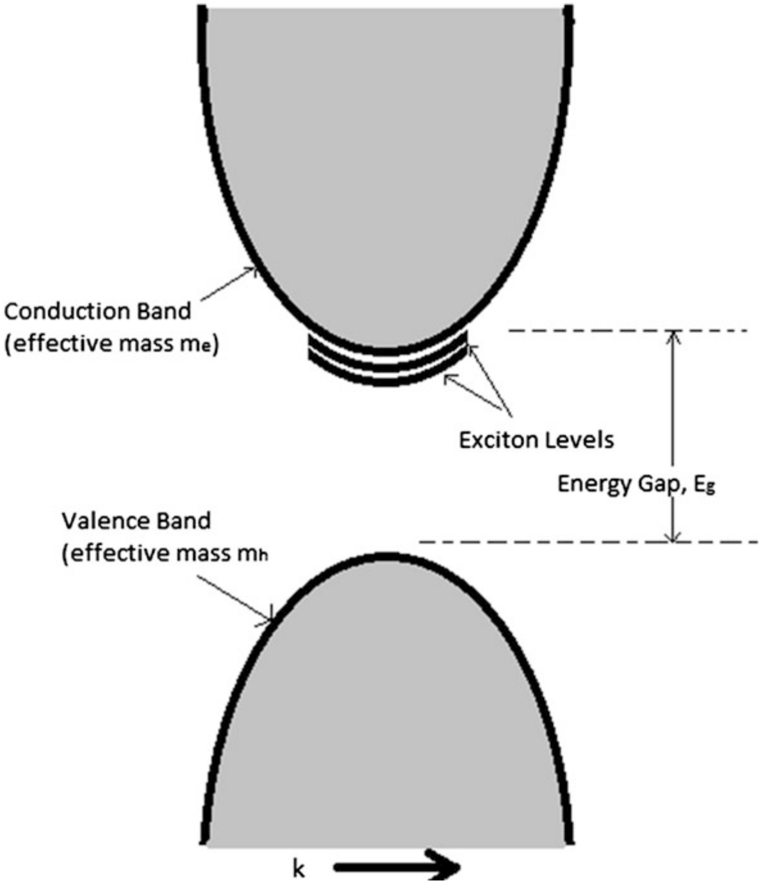


Fig. 6 Formation of excitonic states

constant of the material. Such electron–hole pair system, found mainly within nonmetallic system, is known as exciton. The principle effect of the formation of exciton is that they lower the band gap of the semiconducting material. This type of band gap is called excitonic band gap (shown in Fig. 6). The character and properties of the exciton strongly depends on the nature of the materials in which they form.

According to the formation mechanism, excitons are classified into three groups:

(a) **Wannier–Mott exciton**: Within materials with relatively high dielectric constant, the Coulomb attraction between electron and hole is effectively screened and as a result the exciton is found to be loosely bound and they are delocalized over many lattice sites. This type of exciton is known as Wannier–Mott exciton and is well described by effective mass approximation that originates from nearly free electron model. According this model, the exciton is characterized by wave vector k and the Hamiltonian that describes the exciton can be written as

$$H_{\mathrm{exciton}} = \frac{\hbar^2 k^2}{2m_{\mathrm{e,eff}}} + \frac{\hbar^2 \mathrm{k}^2}{2m_{\mathrm{h,eff}}} - \frac{e^2}{4\pi\varepsilon|r_{\mathrm{e}} - r_{\mathrm{h}}|} \tag{49}$$

where $m_{\mathrm{e,eff}}$ $(m_{\mathrm{h,eff}}b)$ and $r_{\mathrm{e}}(r_{\mathrm{h}})$ represent effective mass of the electron (hole) in conduction (valence) band and its position. ε is the dielectric constant of the material. First two terms represent the kinetic energy of the electron and hole, respectively, and third term illustrates the potential energy acting between electron and hole. This Hamiltonian is analogous to the Hamiltonian of hydrogen atom. If one solves this (solution procedure can be found in any book of quantum mechanics), then binding energy of the electron–hole pair, i.e., exciton (E_x) are characterized by a quantum number (n) similar to hydrogen atom and can be written as

$$E_x = E_n(k) = -\frac{e^2}{2a_{\mathrm{B}}}\frac{1}{4\pi\varepsilon}\frac{1}{n^2} + \frac{\hbar^2 k^2}{2M} \tag{50}$$

where $a_{\mathrm{B}} = 4\pi\varepsilon\frac{\hbar^2}{\mu e^2}$ is the excitonic Bohr radius and illustrates the equilibrium radius corresponding to excitonic states. $M = m_{\mathrm{e,eff}} + m_{\mathrm{h,eff}}$ and $\mu = (m_{\mathrm{e,eff}}^{-1} + m_{\mathrm{h,eff}}^{-1})^{-1}$ represents the total mass and reduced mass of the electron–hole system, respectively. First term signifies the potential energy of the excitonic system whereas second term is related to the kinetic energy of the system. For all practical purpose, second term may be neglected, so the excitonic energy is completely characterized by its potential energy. Thus the excitonic band gap of the semiconducting material, i.e., band gap of the materials in the presence of bound exciton can be written as

$$E_{\text{g,exciton}} = E_{\text{g}} - \frac{e^2}{2a_{\text{B}}}\frac{1}{4\pi\varepsilon}, \quad n = 1 \tag{51}$$

With E_{g} is the band gap of the corresponding material in the absence of exciton. One important point that must be mentioned here that in the presence of exciton, band gap decreases. Among different semiconducting materials around us, ZnO exhibits excitonic effect at room, because excitonic binding energy for ZnO is found to be 60 meV that is stable at room temperature. For other semiconductors, having exciton binding energy less than thermal energy at room temperature, exciton is not stable. So for these materials, no excitonic effect, i.e., band gap lowering due to exciton is observed. It is clear from above equation that the second term that presents binding energy of exciton depends on the Bohr radius (a_{B}). Now if we have nanoparticle having size (R) less than a_{B}, i.e., $R < a_{\text{B}}$, then the excitonic band gap may be written as

$$E_{\text{g,exciton,nanoparticle}} = E_{\text{g,nanoparticle}} - \frac{e^2}{2R}\frac{1}{4\pi\varepsilon} \tag{52}$$

Thus by varying the radius of the nanoparticle, it is possible to tune excitonic energy band gap. Considering size effect, the excitonic band gap of a nanoparticle may be defined, in general, in the following manner

$$E_{\text{g,exciton,nanoparticle}} = E_{\text{g,bulk}} + E_{\text{q}} - \frac{e^2}{2R}\frac{1}{4\pi\varepsilon} \tag{53}$$

where $E_{\text{g,bulk}}$ is the band gap energy of bulk system. E_{q} represents the enhancement of the band gap due to quantum effect (Eq. 48). More detailed calculations on the excitonic energy gap of the nanoparticles considering first excited state of electron

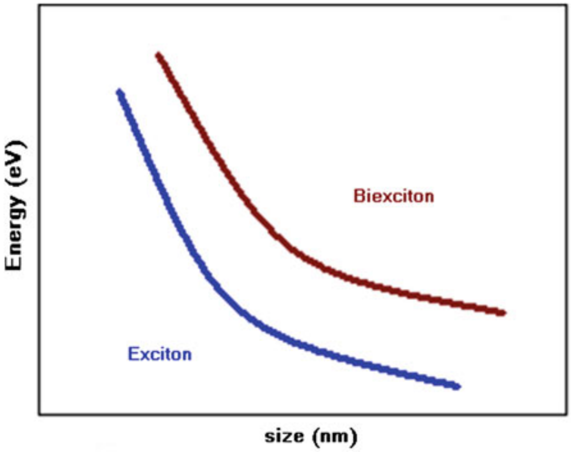


Fig. 7 Variation of exciton and biexciton binding energy with size of the quantum dot

and hole (the quantum state is defined by $1S_e1S_h$), known as Brus model, shows that the above equation may be written as

$$E_{\mathrm{g,exciton,nanoparticle}} = E_{\mathrm{g,bulk}} + \frac{\pi^2\hbar^2}{2R^2}\left(\frac{1}{m_{\mathrm{e,eff}}} + \frac{1}{m_{\mathrm{h,eff}}}\right) - \frac{e^2}{2R}\frac{1.786}{4\pi\varepsilon} \tag{54}$$

The variation of excitonic binding energy with respect to particle size is shown in Fig. 7.

(b) **Frenkel exciton**: In the case of material with low dielectric constant, the Columbic force between electron and hole is very strong. In such systems, they are very much localized to a particular atom. This type of exciton is called Frenkel exciton. The properties of the Frenkel exciton are described by tight-binding model of band structure.

According to this model, the ground state wave function (Ψ_G) of N identical molecules can be written as the product of their individual wave function (ψ_i),

$$\Psi_G = \prod_{n=1}^{N} \psi_n \tag{55}$$

and excited state wave function, when one of the atom is excited (let "a" atom) can be written as

$$\phi_a = \psi_a \prod_{\substack{n=1 \\ n \neq a}}^{N} \psi_n \tag{56}$$

Since, any atom within the system may be excited, so the wave function of the excited state may be superposition of them and the general wave function of the excited state may be written as

$$\Psi_E(k) = \sum_{a=1}^{N} C_a(k)\phi_a \tag{57}$$

where "k" is the wave vector that represents kth excited state. Using periodic boundary and normalization condition for wave function similar to tight-binding model, $\Psi_E(k)$ may be written as

$$\Psi_E(k) = \frac{1}{\sqrt{N}}\sum_{a=1}^{N} \exp\left(\frac{2\pi ika}{N}\right)\phi_a \tag{58}$$

The above equation illustrates the superposition of the excited states of the atoms, i.e., excitation is delocalized over N atoms. In this context, typical length, known as coherent length, is defined as the as the length determined by the array of N atoms. In this excitonic process, the absorption and emission process involves large number of atoms; hence the strength of the optical transition (also known as oscillator strength) is given by the sum of the strength of optical transition of the each atom, present within this length. For system having dimension less than this characteristic length, optical strength is proportional to the number of atoms present there. Another important consequence of this exciton is the radiative decay rate that is inversely proportional to optical strength. Since, optical strength decreases with size of the nanoparticles, so excitonic life time is found to be increased with decreasing number of atom. Such phenomenon is called as “superradiance” by analogy with the similar phenomena occurring in atomic and molecular ensembles.

(c) Charge transfer exciton: This type of exciton is mainly found in ionic crystal or molecular crystal having two kinds of molecule with different ionization potential and electron affinity. Here electron and hole are tightly bound and delocalized on neighboring atoms. It is an intermediate exciton between Frenkel and Wannier–Mott excitons.

Another important quasiparticle, found in nanostructure materials in recent time, is the biexciton. This newly formed quasiparticle is created in case of nonzero binding energy that is defined as

$$\Delta E_{XX}^{\mathrm{B}} = 2E_X - E_{XX}(K) \tag{59}$$

With $E_{XX}(K)$ is the biexciton binding energy. Physically, when more than one electron is excited, electron–electron, electron–hole, and hole–hole interactions have to consider to calculate the binding energy of the biexciton. The dispersion relation for biexciton may be written as

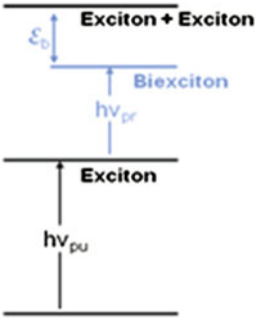


Fig. 8 Biexciton formation in pump-probe experiment

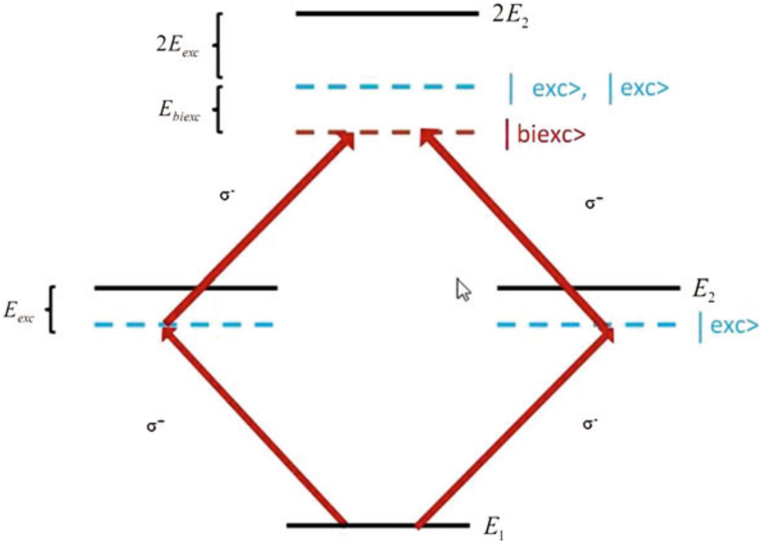


Fig. 9 Diamond structure for biexciton transition scheme

$$E_{XX}(K) = 2E_X - \Delta E_{XX}^{\mathrm{B}} + E_{\mathrm{exch}} + \frac{\hbar^2 K_{XX}{}^2}{4M_X} \tag{60}$$

where K_{XX} presents the wave vector, presenting the motion of the center of mass, and signifies the kinetic energy of the system. E_{exch} represents electron–hole exchange interaction energy. In general, the formation of biexciton is observed either in pump-probe experiments that involves one pump-photon and one probe-photon (shown in Fig. 8) or in two-photon absorption process.

The biexciton ground state is characterized by two electrons and two holes with antiparallel spin such that resultant angular momentum of the system is zero. Experimentally, biexciton binding energy strongly depends on the effective mass of electrons and holes within the quantum structure. Size and shape of the quantum particle strongly affects the binding energy of biexciton and associated optical properties. Generally, it increases with size reduction. In this point, it has to be mentioned that the elementary transitions to excited two-photon states is forbidden by quantum mechanical selection rule, but in the case of quantum particles, the transition becomes possible due to relaxation of the selection rule caused by Coulomb potential. The biexciton transition is understood using so-called diamond scheme (shown in Fig. 9) that illustrates the selection rule in the presence of circularly polarized light. A photon with σ^+ polarization excites spin-up electron and spin-down is created by σ^- photon.

4 Magnetic Property

Materials having response in the presence of an applied magnetic field are called magnetic materials. They have either intrinsic magnetic moment or magnetic moment induced by magnetic field. In the presence of an applied magnetic field, moments will be aligned in the direction of applied field and will setup its own field that will modify the applied magnetic field. Magnetic moments are regarded as microscopic currents and they act as a source of magnetic induction (**B**). In fact we may write

$$\mathbf{B} = \mu_0(\mathbf{M} + \mathbf{H}) \tag{61}$$

where **M** and **H** represent the magnetic moment per unit volume, called magnetization, and magnetic field strength associated with macroscopic current, respectively. **M**, having the same dimension as **H**, empirically is found to be proportional to **H**, i.e., $\mathbf{M} = \chi\mathbf{H}$. The proportionality constant is called magnetic susceptibility and is a dimensionless quantity. The primary objective of studying magnetic properties is to determine χ for them and to investigate its dependence on temperature (T) and **H**. Substituting for **H** into Eq. (1), we may write **M** in terms of **B** and χ,

$$\mathbf{M} = \frac{\chi\mathbf{B}}{\mu_0(1+\chi)} \tag{62}$$

From Eq. (62), it is clear that for materials having very small value of $\chi(1 \gg X), \mathbf{M} \simeq \frac{\chi\mathbf{B}}{\mu_0}$ and their magnetization contribution to **B** is very small. These special classes of materials possesses magnetic moments only in the presence of magnetic field and are classified into two categories according to the sign of χ. Materials having $\chi < 0$ are called diamagnetic materials and materials with $\chi > 0$ are called paramagnetic materials. The significance of negative χ is that **M** lies in the opposite direction of **H**. The origin of diamagnetism is the magnetic dipole moment induced by applied magnetic field according to Lenz's law. In paramagnetic materials, constituting atoms have permanent noninteracting magnetic moments due to their intrinsic spin and orbital motion of the unpaired electrons. In the absence of magnetic field, they are thermally randomized and give zero magnetic moments. In the presence of magnetic field, they are aligned in the direction of field against thermal randomization and χ for paramagnetic materials is found to be dependent on temperature following Curie law $\chi = \frac{C}{T}$. C, a positive constant, is known as Curie constant. From slope of the experimentally measured χ versus T^{-1} curve, it is possible to find out the value of C. After knowing C, one may calculate magnetic moment of the constituting atoms. Field dependence of magnetization of these paramagnetic materials instead of classical Langevin function is given by Brillouin function $B_j(a)$,

$$M = M_{\mathrm{s}} B_j(a) = M_{\mathrm{s}} \left[\left(\frac{2j+1}{2j} \right) \coth\left(j \cdot a + \frac{a}{2} \right) - \frac{1}{2j} \coth\left(\frac{a}{2} \right) \right] \qquad (63)$$

with $a = \frac{g\mu_{\mathrm{B}} H}{k_{\mathrm{B}} T}$, where "*g*" and μ_{B} are the Landé *g*-factor Bohr magnetron, respectively, "*j*" represents the quantum number for the total (spin and orbital) angular momentum of the constituting atoms. For example, *j* is equal to 5/2 for Mn^{+2} ions.

There are other classes of materials with $\chi \gg 1$ due to interacting unpaired electrons of their constituting atoms. One of them is ferromagnetic materials (Fe, Co, Ni) having spontaneous magnetization within small regions even in the absence of any external magnetic field. These regions each having large number of atoms with interacting unpaired electrons and having net magnetization along some direction are called domains. Domains are distributed randomly in the absence of magnetic field and give zero net magnetization. When they are exposed to magnetic field, they are aligned in the direction field and result nonzero magnetization. Spontaneous magnetization corresponding to each domains is maximum at $T = 0$ K and drops to zero at a critical temperature, known as ferromagnetic Curie temperature (T_{c}). Ferromagnetism is supposed to be a collective phenomenon since the mutual interaction among atomic moments promotes alignment with one another and is responsible for temperature and field dependent behavior of magnetization. There exists two different theories namely mean field theory and Heisenberg theory to explain magnetic behavior of ferromagnetic materials. According to mean field theory, there exists a nonlocal internal magnetic field, known as Weiss field, which acts on the magnetic moments to align in the absence of external magnetic field (H_{int}). H_{int} is assumed to be proportional to magnetization of the ferromagnetic domains, i.e., $\mathbf{H_{int}} = \theta_W \mathbf{M}$, where θ_W, the proportionality constant, is known as Weiss molecular field constant. Thus in the case of ferromagnetic materials, total field acting on each magnetic moment is sum of $\mathbf{H_{int}}$ and $\mathbf{H}$. So from analogy with paramagnetic materials, the average magnetization of the sample can be written as

$$M = M_{\mathrm{s}} B_j(a), \quad \text{with } a = \frac{g\mu_{\mathrm{B}}(H + H_{\mathrm{int}})}{k_{\mathrm{B}} T} \qquad (64)$$

Under appropriate situations, solution of Eq. (4) will give us nonzero magnetization, i.e., the system exhibits spontaneous magnetization at $H = 0$. For $T > T_{\mathrm{c}}$, the only solution corresponding to Eq. (64) is $M = 0$ when $H = 0$, i.e., the sample does not possesses spontaneous magnetization in the absence of magnetic field. Physically, we can say that below T_{c}, alignment of moments is established due to internal magnetic field and gives nonzero magnetization, but above T_{c}, the alignment is destroyed by thermal energy that results zero spontaneous magnetization and their behavior becomes paramagnetic.

The second theory was developed later by Heisenberg on purely quantum mechanical basis in the year 1928. According to him, the spontaneous magnetization results from exchange interaction between neighboring magnetic moments. Within this theory, it is assumed that the neighboring atomic moments (due to spin) are coupled via exchange interaction that gives rise to exchange potential energy of the material. The quantum mechanical Hamiltonian of such system can be written in the following form

$$H = -\sum_{i>j} J_{ij} s_i s_j \tag{65}$$

where $s_i = \pm 1$ is the ith spin and J_{ij} is the exchange interaction between ith and jth spins. Positive J_{ij} favors the parallel alignment of the neighboring spins and results ferromagnetism. The Heisenberg exchange interaction just replaces θ_W to explain the temperature dependence of the magnetization.

Since ferromagnetic materials above T_c is found to behave paramagnetically, their temperature dependence of χ above $T_{c,}$ is given by the following relation $\chi = \frac{C}{T - T_c}$, known as modified Curie law or Curie–Weiss law. The magnitude of T_c depends on the strength of the exchange interaction among magnetic moments. An empirical method, based on electron–electron Coulomb interaction and Pauli exclusion principle, was developed to calculate the exchange interaction in the case of solid constituting of transitional metal. In these systems, the ratio of the atomic separation (a) to the diameter of 3-d orbitals (r) corresponding to transitional metals was used to describe the overlap between nearest neighbor d-orbitals. Figure 10 shows the variation of exchange interaction with D/d and the curve is known as Bethe-Slater curve, after their names. To make the problem more simple, J_{ij} was approximated to be constant ($J_{ij} = J_{ex}$) and independent of atomic sites, then Curie temperature is found to be related to J_{ex} as follows:

$$T_c = \frac{2ZJ_{ex}S(S+1)}{3k_B} \tag{66}$$

From curve, it is clear that for some transition metal atoms, like Co, Ni, Gd, the exchange interaction is positive, i.e., neighboring moments are parallel and they possesses ferromagnetic property. On the other hand, there also exist some

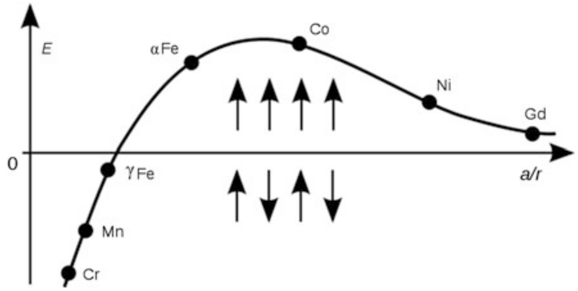


Fig. 10 Bethe–Slater curve showing variation of exchange interaction with a/r ratio (http://commons.wikimedia.org/wiki/File:Bethe-Slater_curve_by_Zureks.svg)

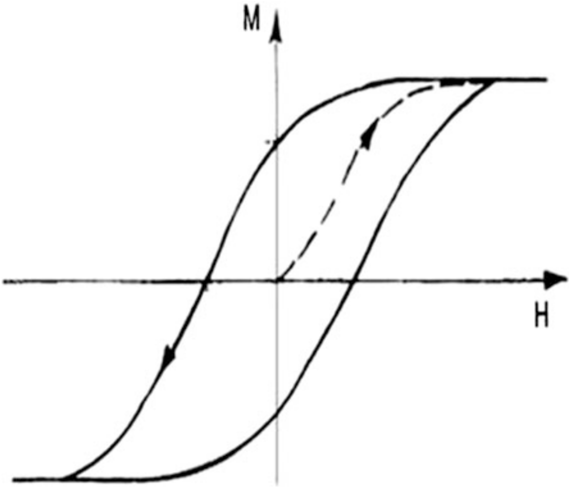


Fig. 11 Typical hysteresis loop of ferromagnetic materials

materials such as Mn, Cr with negative exchange interaction that favors antiparallel alignment of neighboring moments. These kinds of materials are known as antiferromagnetic materials. The magnetic behavior of antiferromagnetic materials is modeled on the basis of two interpenetrating sublattices, one of which has its magnetic moment pointing up and the other one down. Similar to ferromagnetic materials, antiferromagnetic materials also exhibits a characteristic temperature, known as Neel temperature (T_N), below which magnetic moments have long-range order, but above which they become paramagnetic and follows Curie–Weiss law for antiferromagnetic materials $\chi = \frac{c}{T - T_N}$. The net magnetic moment of antiferromagnetic materials is found to have relatively small compare to ferromagnetic materials. Figure 11 shows one of the typical $M(H)$ curve, known as hysteresis curve, that is obtained on cycling the external field.

In this context, three parameters are defined; (a) remanent magnetization (M_r) that signifies the residual magnetization after removing the magnetic field, (b) coercive field (H_c) that is defined as the minimum reverse field required to obtain magnetization to zero and saturated magnetization (M_s) which illustrates the maximum magnetization, obtained from the sample. It is worthy to mention that all these parameters are found to be temperature dependent. Technically, the hysteresis loop illustrates magnetic property of the ferromagnetic materials and its shape is determined by the minimization of magnetic free energy of the system that involves field energy (Zeeman energy), self-field (demagnetization energy), wall energy, and magnetic anisotropy energy. The general form of the magnetic field energy in the presence of magnetic field **H** (sum of external magnetic field and demagnetizing field) can be written as follows [7]:

$$E = \int \left[A(\mathbf{r}) \left[\frac{\nabla \mathbf{M}}{\mathbf{M_s}} \right]^2 - K(\mathbf{r}) \left[\frac{\mathbf{M} \times \mathbf{n}}{\mathbf{M_s}} \right]^2 - \mu_0 \mathbf{M} \times \mathbf{H} \right] \mathrm{d}\mathbf{r} \tag{67}$$

$A(\mathbf{r})$ is called the local exchange stiffness ($A = J_{\text{ex}}S^2/a$ at 0 K, and "a" represents the interatomic spacing). $K(\mathbf{r})$ represents the local magnetic anisotropy energy density, **n** is the unit vector parallel to the easy direction of magnetization. Magnetic anisotropy energy illustrates the angular dependence of magnetic energy, i.e., dependence of magnetic energy on angles θ and φ between magnetization direction and easy axis of magnetization. Physically, it acts as a barrier for switching the magnetization. In case of real materials, it is subdivided into few categories such as magnetocrystalline anisotropy, shape and stress anisotropies. Coercive field (H_c) in the hysteresis curve depends on the anisotropy energy of the materials. Stress anisotropy originates from magnetostriction and is defined as the process by which shape of the ferromagnetic materials changes during magnetization as a result magnetoelastic effect. Shape anisotropy is associated with the demagnetization effects and is minimized by producing domain with high aspect ratio. In the case of various nanostructures, shape anisotropy plays a crucial role during demagnetization processes; hence H_c may be tuned by varying shape of the nanostructure. Magnetocrystalline anisotropy results from coupling of the magnetization and the easy axis of crystalline. Physical origin of such anisotropy is the spin–orbit coupling.

As it was mentioned that the characteristic feature of ferromagnetic materials is the domain structure within which moments are aligned parallel due to long-range order and possesses large moments. In the absence of magnetic field, because of vector cancelation of the magnetization corresponding to each domain, net magnetization is zero. Domains are fully aligned in the presence of sufficiently large magnetic field and attain M_s. Main mechanism behind the enhancement of the magnetization with increasing magnetic field is as follows: as the external magnetic field is increased, nucleation and rotation of the domain in the direction field occurs that increases magnetization in the direction of field. After determining M_s experimentally, it is possible to calculate magnetic moment of the constituting atoms. For ferromagnetic material having more than one domain structure, an interfacial region between domains that illustrates the variation of magnetization from one domain to another domain exists. This is called domain wall. Size of the domain wall is determined by the volume, magnetostatic and interfacial domain wall energies. In general, the domain wall energy (γ_{wall}) can be written as the sum of exchange (γ_{exc}) and anisotropy energy (γ_K) terms, given by [8]

$$\gamma_{\text{wall}} = \gamma_{\text{exc}} + \gamma_K \tag{68}$$

In general γ_K is expressed by the relation $\gamma_K = KNa$, with K is anisotropy energy constant, "a" is interplanar spacing and N represents the number of layer involved in the rotation process. Thus Na is the width of the domain wall. γ_{exc} is given by the equation $\gamma_{\text{exc}} = \frac{\pi^2 J_{\text{ex}} S^2}{\text{Na}^2}$. The equilibrium number of planes (N_{eq}), involved in the domain wall, can be obtained using the condition $\frac{\partial \gamma_{\mathrm{wall}}}{\partial N} = 0$. It is found to be $N_{\mathrm{eq}} = \sqrt{\frac{\pi^2 J_{\mathrm{ex}} S^2}{Ka^3}}$, and subsequently width of the domain wall is $\sqrt{\frac{\pi^2 J_{\mathrm{ex}} S^2}{Ka}}$. The other length scale that is defined as the distance over which the perturbation due to the switching of a single spin decays, known as ferromagnetic exchange length (L_{ex}), i.e., domain size, and its expression is given by $L_{\mathrm{ex}} = \frac{36\sqrt{AK}}{\mu_0 M_s^2}$ [9] in the absence of any magnetic field. In order to differentiate between hard and soft magnetic material, the dimensionless parameter (ξ), called magnetic hardness parameter, is defined by $\xi = \frac{\pi L_{\mathrm{ex}}}{Na}$. ξ is order of unity in case of hard material, i.e., $L_{\mathrm{ex}} \sim Na$. On the other hand, ξ is found to be very much less than one in case of soft material. Temperature dependence of magnetization corresponding to these magnetic materials is investigated under FC–ZFC protocol. This is carried out by first annealing the sample at a certain temperature where its spontaneous magnetization vanishes and followed by cooling the sample in the absence of any magnetic field. The magnetization data is recorded during heating cycle in the presence of magnetic field. This step is known as zero-field-cooled (ZFC). On field cooling (FC) protocol, the sample is cooled in the presence of magnetic field (same filed as ZFC) and other steps are same as ZFC. One typical set of FC–ZFC curve is shown in Fig. 12. The curve reveals that there exists a typical temperature, called irreversibility temperature (T_{irr}), above which both FC and ZFC merge with each other, but at below T_{irr}, there exists a difference between them. Reduction of ZFC magnetization with decreasing temperature is related to disordered states that are trapped in the ZFC measurement. In this temperature range, magnetization starts off from zero only from contribution those are not related to aligned domains. Up to T_{irr}, ZFC magnetization originates from nonequilibrium state and their difference depends on the magnetic anisotropy energy of the materials.

The above-mentioned Bethe–Slater curve well explains the magnetic property of transitional metallic system. The exchange interaction acts between nearest neighbors. This is the reason why it is also called direct exchange. But there also exists other magnetic system like Fe_3O_4, $ZnFe_2O_4$, $NiFe_2O_4$, diluted magnetic

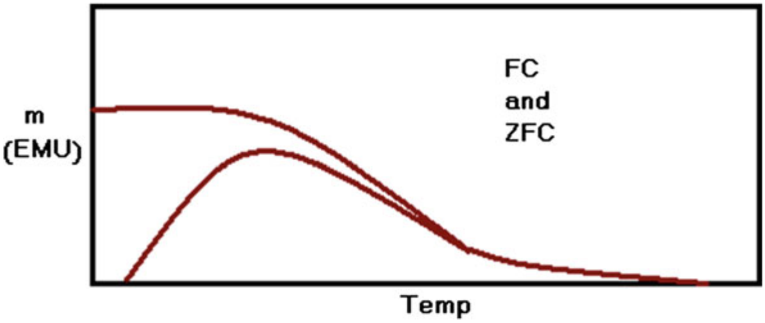


Fig. 12 FC–ZFC curve of typical ferromagnetic material

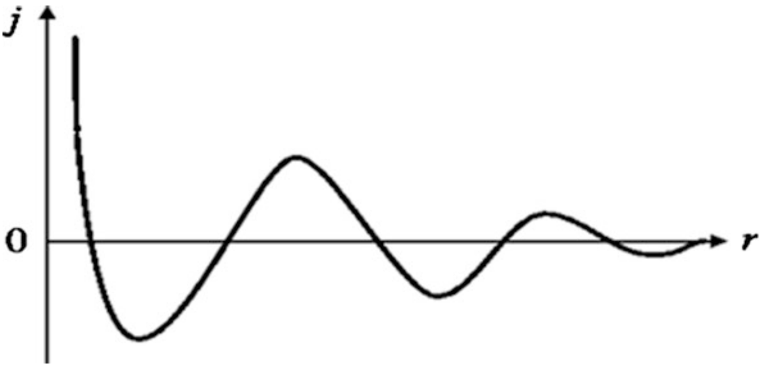


Fig. 13 The coefficient of indirect (RKKY) exchange versus the interatomic

semiconductor, etc., where magnetic ions are separated by some distance such that the direct interaction between them is not possible. In order to explain their magnetic behavior, concept of different types of exchange interaction like indirect exchange, superexchange, etc., were developed.

Indirect exchange acts between moments with relatively large distance where the direct exchange interaction is negligible low. This type of exchange interaction is basically RKKY type that can be either positive or negative depending on the separation between them. The basic mechanism, involved in this process, is that magnetic ion causes a spin polarization in the conduction electrons in its neighborhood. This spin polarization of the conduction electron is felt by another moments within the range. So in this way two moments, located at different position, are coupled, the coupling is known as indirect coupling. Figure 13 shows the variation of exchange interaction between moments with their in between distance. Magnetic property of diluted magnetic semiconducting materials is explained on the basis of this indirect exchange interaction.

There also exists another type of indirect exchange, known as superexchange, which acts between magnetic ions, separated by nonmagnetic anions. For example, the exchange interaction between two Fe ions, separated by oxygen ions within iron oxides, is found to be superexchange. Nature of this exchange interaction is found to be negative when magnetic ions (on each side of the anion) carry equal magnetic moments.

One of the significant effects of size on the magnetic property is the reduction of M_s. M_s is observed to be decreased significantly with decreasing particle size. The broken exchange bonds and breaking of translational symmetry of the lattice at the surface of the nanoparticles induce disorder in the spin arrangement near surface that leads zero magnetization in the low-field region, whereas it possesses paramagnetic contribution in the high-field region. The increase of surface contribution with decreasing particle size causes reduction of M_s. Let "D" be particle size and "t" be the thickness of such magnetically disordered layer. Then from simply considering of volume responsible for magnetization, we may write the following relation between saturation magnetization of nanoparticles ($M_{\mathrm{s,nano}}$) and that of bulk ($M_{\mathrm{s,bulk}}$) [10]

$$M_{\mathrm{s,nano}} = M_{\mathrm{s,bulk}} \frac{\frac{4}{3}\pi\left(\frac{D-2t}{2}\right)^3}{\frac{4}{3}\pi\left(\frac{D}{2}\right)^3} = M_{\mathrm{s,bulk}}\left(1 - \frac{6t}{D}\right) \tag{69}$$

Variation of M_{s} as given by Eq. (69) is well supported by many experimental results. The other consequence of size effect on the magnetic property arises in the form of the temperature dependence of M_{s}. We know that within crystalline materials, atoms are periodically arranged. As we increase the temperature, they vibrate collectively after losing their individual identity and will generate vibrational wave. Quantum mechanically their vibrational energy is quantized and quantum unit of lattice vibrational energy is called phonon. Vibration mode is characterized by wavenumber (k) and its associated phonon energy is $E_{\mathrm{lattice},k} = \omega k^2$ for low value of k, ω represents the velocity of vibrational wave. Similarly to it, we may consider ferromagnetic or ferrimagnetic materials as periodic arrangement of spin. It is a common experience that if we increase the temperature of magnetic system, its magnetization will decrease. Thus the demagnetization effect may be thought as generation of some spin wave. Similar to lattice vibration, we may define the quantum of the spin wave energy as magnon that will be characterized by a wavenumber (k). For low value of k, i.e., for higher value of the corresponding wavelength, we may write the dispersion relation between magnon energy and wavenumber, given by

$$E_{\mathrm{magnon},k} = Dk^2 \tag{70}$$

where "D" is called spin wave stiffness coefficient. Then the reduction of M_{s} due to increasing temperature from its value at $T = 0$ K can be written as [11]

$$M_{\mathrm{s}}(T) = M_{\mathrm{s}}(0) - \int_0^\infty g(k)n(k)\mathrm{d}k \tag{71}$$

where $g(k)$ represents the density of spin wave and $n(k)$ is the Bose–Einstein distribution function that gives occupation probability of any states for magnons. Using exact form of $n(k)$, one may obtain,

$$\frac{M_{\mathrm{s}}(T) - M_{\mathrm{s}}(0)}{M_{\mathrm{s}}(0)} = \frac{g\mu_{\mathrm{B}}}{2\pi^2}\int_0^\infty \frac{k^2\mathrm{d}k}{e^{\frac{E_{\mathrm{magnon,k}}}{k_{\mathrm{B}}T}} - 1} \tag{72}$$

In case of bulk ferromagnetic system, we may assume spin wave states continuous, and then the above equation is simplified into the following form [12]

$$M_{\mathrm{s}}(T) = M_{\mathrm{s}}(0)[1 - BT^{\frac{3}{2}}] \tag{73}$$

with $B = 2.6\ V_0\left(\frac{k_{\mathrm{B}}}{4\pi D}\right)^{\frac{3}{2}}$, V_0 is the atomic volume. This equation is the famous Bloch's law for temperature dependence of M_{s} corresponding to bulk ferromagnetic materials. In the case of nanoparticle, two interesting situations arise. First, due to reduced size of the system, spin-density states are no longer continuous. We may quantize the energy by simply considering cubic shape of the particle with size L that is given by

$$E_{\mathrm{magnon},n} = D\left(\frac{n\pi}{L}\right)^2, \quad n = 1, 2, 3\ldots \tag{74}$$

Second, low energetic magnon having wavelength greater than the particle size will not excited. Under these two assumptions we may write

$$M_{\mathrm{s}}(T) = M_{\mathrm{s}}(0) - \frac{M_{\mathrm{s}}(0)}{NS}\sum_{n=1}^{p}\frac{1}{e^{\frac{E_{\mathrm{magnon,n}}}{k_{\mathrm{B}}T}} - 1} - \int_{k'}^{\infty} g(k)n(k)\mathrm{d}k \tag{75}$$

Summation will run until $E_{\mathrm{magnon},p} \sim k_{\mathrm{B}}T$. For all temperature far below T_{c} of the material, third term may be neglected. Thus the above equation is simplified to

$$M_{\mathrm{s}}(T) = M_{\mathrm{s}}(0) - \frac{M_{\mathrm{s}}(0)}{NS}\left[e^{-\frac{E_1}{k_{\mathrm{B}}T}} + e^{-\frac{E_2}{k_{\mathrm{B}}T}} + e^{-\frac{E_3}{k_{\mathrm{B}}T}} + \cdots\right] \tag{76}$$

Size effect on M_{s} becomes ineffective unless thermal energy is less than the energy gap between two successive spin-states.

Not only exhibiting reduced M_{s} or violation of Bloch's theorem, nanomagnetic materials often possesses completely new type of magnetic response. One of it is the spin-glass response, in which the magnetic interactions are characterized by quenched randomness: a given pair of moments is associated with roughly equal probability of having a ferromagnetic or an antiferromagnetic interaction. It is well known from previous discussion that the exchange interaction between moments strongly depends on their distance. In the case of nanoparticles, large fractions of surface atoms are often associated with broken bond and vacancies that cause variation in the interatomic distances [13]. Since, exchange interaction strongly depends on the interatomic distances, so their exchange interactions are getting frustrated and they exhibits ferromagnetic or antiferromagnetic type of exchange interaction with equal probability. There also happens another interesting situation in nanoparticle systems (ferromagnetic in bulk). We may divide the nanoparticle into two regions; (a) core region with periodicity same as bulk and (b) shell region with aperiodic arrangements of atom-near surface. Then moments in the core region show ferromagnetic alignment, whereas moments in the shell region correspond either antiferromagnetic or spin-glass behavior. Hysteresis loop of such core–shell structure is being shifted from origin. This effect is known as exchange bias effect and is useful for magnetic storage devices.

Ferromagnetic nanoparticles exhibit completely new type of magnetic response when their size becomes comparable or less than the domain size. In this range of particle size, they are broadly classified into two categories, namely single domain ferromagnetic particles and superparamagnetic particles. Like bulk ferromagnetic particles, an array of single domain magnetic nanoparticles also possesses hysteresis loop that is associated with M_{s}, M_{r}, and H_{c}. But the main difference lies in the mechanism by which the magnetization is cycled through the hysteresis loop. The traditional method by which bulk ferromagnetic material experiences the enhancement of magnetization with increasing field is nucleation and rotation of domain wall movement as well as rotation of the magnetization away from easy axis of magnetization. But in the case of array of single domain particle, the mechanism involves only coherent rotation of magnetization. Other interesting behavior of ferromagnetic materials having size less than the critical domain size is their paramagnetic response. All spins within each domain are aligned in a particular direction at low temperature due to strong exchange interaction and the domain possesses large magnetic moment. Such single domain particles are called "superparamagnetic particles." For magnetization reversal, they have to overcome the anisotropy energy barrier (E_{a}) from one stable configuration to other. The energy is determined by size and shape of the particles and in the case of spherical nanoparticles with volume V it is given by $E_a = KV\sin^2\theta$, where K is the anisotropy constant and θ represents the angle between easy magnetization axis and magnetization [14]. Superparamagnetic particles possesses ferromagnetic or ferrimagnetic behavior at low temperature that is realized by the hysteresis loop in the magnetization curve. As temperature is increased, there exists a critical temperature, known as superparamagnetic blocking temperature (T_{B}), above which thermal energy is sufficient to overcome the anisotropy energy. At $T > T_{\mathrm{B}}$, magnetization is no longer stable and magnetic response of the particle is paramagnetic like. They are characterized by reversible magnetization without hysteresis due to coherent rotation of the spins within each single domain. In contrast to quantized behavior of atomic moments in paramagnetic materials, superparamagnetic particles behave classically, i.e., they can make any angle with external magnetic field. So their response to external magnetic field instead of Brillouin function (Eq. 3) is given by Langevin function

$$M = N\mu\left[\coth\left(\frac{3g\mu_{\mathrm{B}}H}{k_{\mathrm{B}}T}\right) - \frac{k_{\mathrm{B}}T}{g\mu_{\mathrm{B}}H}\right] \tag{77}$$

The other important parameter, associated with superparamagnetic behavior of single domain particles, is the relaxation time of spins (τ) due to thermal fluctuation and it is given by

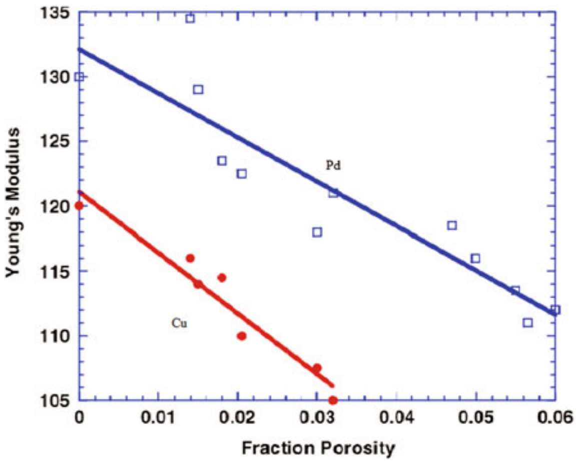


Fig. 14 Young's modulus as a function of porosity for nanocrystalline Pd and Cu (Reprinted with permission from Meyers et al. [20] © Elsevier 2006)

$$\tau = \tau_0 \exp\left(\frac{KV}{k_B T}\right) \tag{78}$$

where K is the anisotropy constant, V is the volume of the nanoparticle, thus KV represents the anisotropy energy of the superparamagnetic particle. Preexponential factor (τ_0) is constant and its value was estimated to be 10^{-9} s by Néel. Derivation of this equation is given elsewhere [15]. Superparamagnetism particles are found to have many technological applications in the field of drug delivery, hyperthermic therapy for treatment of cancer cells, etc.

5 Mechanical Property

Over the years, the versatile properties of the nanoparticles that is being introduced due the presence of large fraction of surface atoms open op many new application opportunities. Many nanodevices like nanoreactor, nanogenerator, etc., have been developed based upon nanomaterials. As result, investigation of their different mechanical properties such as ductility, strain hardening, yield stress, dynamic response, creep, etc., becomes essential. Processing flaws like porosity are known to have significant effect on the mechanical properties of the nanomaterials. Young's modulus (Y) of nanocrystalline Pd and Cu, measured by Weertman [16] was found to be decreased with porosity (p) (Fig. 14). This linear variation, according to Wachtman and Mac Kenzie [17] can be written as

$$Y = Y_0\left(1 - Ap + Bp^2\right) \tag{79}$$

with Y_0 is the Young modulus corresponding bulk system and A, B are constants.

Grain size also modifies yield stress (σ). The purely empirical relation between grain size (d) and stress (σ_d) is given by Hall-Petch equation,

$$\sigma_d = \sigma + kd^{-\frac{1}{2}} \tag{80}$$

Another important mechanical property that is of interest in recent time is the ductility. It is found to be increased as its size is reduced. In is worthy to mention that the metallic nanoparticles possesses smallest ductility for grain size <25 nm (77). Normally, there are three different contributions to enhance ductility of the nanocrystalline materials: (a) artifacts, i.e., pores, (b) tensile instability and (c) crack nucleation or shear instability.

6 Dielectric Property

For most of the materials, dielectric property is independent of the strength of the electric field, but it depends on the frequency in the case of alternating field. Oxide-based semiconducting materials like ZnO, TiO_2 are being widely used to fabricate optical devices. The optical usage relies on refractive constant (n), extinction coefficient (k) of the materials. Since n and k are related to dielectric constant (ε) of the materials by the following relation $n + ik = \sqrt{\varepsilon}$, so the knowledge of ε becomes essential. In general, dielectric constant of the material is complex in nature ($\varepsilon = \varepsilon_1 + i\varepsilon_2$) and depends on the frequency of the external electric field that we are using. Simplest theory to describe ε is the Drude–Lorentz–Sommerfeld theory that has been developed. According to this theory, electrons, near conduction band edge, contribute to "n" or "k", hence to ε. Considering classical theory for these free electrons, we may write their equation of motion in the presence of an external electric field $\boldsymbol{E} = \boldsymbol{E}_0\exp(-i\omega t)$,

$$m_{\mathrm{e,eff}}\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} + m_{\mathrm{e,eff}}\Gamma\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = -e\boldsymbol{E} \tag{81}$$

where Γ represents the phenomenological damping constant. Since, mass of the electron is very low, so they can follow the frequency external electric field. We may write the frequency dependent position of them $\boldsymbol{r} = \boldsymbol{r}_0\exp(-i\omega t)$. Substituting this into the above expression,

$$-m_{\mathrm{e,eff}}\omega^2\boldsymbol{r}_0 - i\omega m_{\mathrm{e,eff}}\Gamma\boldsymbol{r}_0 = -e\boldsymbol{E}_0 \tag{82}$$

Solving the above equation for $\boldsymbol{r}_0$, we may obtain $\boldsymbol{r}_0 = \frac{e\boldsymbol{E}_0}{m_{\mathrm{e,eff}}(\omega^2 + i\omega\Gamma)}$ $\boldsymbol{r}_0$ represents the displacement of the electron in the presence of electric field. If "n" be the number of electron per unit volume of the system, then $-ne\boldsymbol{r}_0$ will give the electric dipolemoment generated per unit volume, i.e., dielectric polarization (**P**), so it is given by

$$\mathbf{P} = -ne\mathbf{r}_0 = -\frac{ne^2\mathbf{E}_0}{m_{\mathrm{e,eff}}(\omega^2 + i\omega\Gamma)} \tag{83}$$

Considering the relation among, **P**, **E** and ε that is given by $\varepsilon = 1 + \frac{\boldsymbol{P}}{\varepsilon_0 \boldsymbol{E}}$, we may write the following expression for dielectric constant of the material,

$$\varepsilon = 1 - \frac{ne^2}{\varepsilon_0 m_{\mathrm{e,eff}}(\omega^2 + i\omega\Gamma)} = 1 - \frac{\omega_{\mathrm{p}}^2}{(\omega^2 + i\omega\Gamma)} \tag{84}$$

with $\omega_{\mathrm{p}}^2 = \frac{ne^2}{\varepsilon_0 m_{\mathrm{e,eff}}}$. ω_{p} is called plasma frequency and physically it represents the collective oscillation of wave. For small damping $(\omega \gg \Gamma)$, the real and imaginary part of ε are given by

$$\varepsilon_1 \approx 1 - \frac{\omega_{\mathrm{p}}^2}{\omega^2} \quad \text{and} \quad \varepsilon_2 \approx \frac{\omega_{\mathrm{p}}^2}{\omega^3}\Gamma \tag{85}$$

For any free electron within bulk system, moving with Fermi velocity (v_{F}) and having mean free path “L,” the damping constant may be written as $\Gamma = \frac{\vec{v}_{\mathrm{F}}}{L}$. Physically, the origin of the damping is the scattering of electron from impurity, lattice imperfection, etc.

Now for any electron within nanostructure, it will be scattered from surface of the nanoparticle (with diameter D), since at the surface, lattice periodicity is lost due to broken bond. So in addition to interior imperfection contribution ($\Gamma_{\mathrm{i}} = \frac{v_{\mathrm{F}}}{L}$), there will be another contribution to damping and we may write this contribution as $\Gamma_{\mathrm{s}} = \frac{Av_{\mathrm{F}}}{D}$, A is the proportionality constant. Considering these two contribution, we may the expression for imaginary part of the dielectric constant,

$$\varepsilon_2 = \frac{\omega_{\mathrm{p}}^2}{\omega^3}\left(\Gamma_{\mathrm{i}} + \frac{Av_{\mathrm{F}}}{D}\right) \tag{86}$$

Therefore, ε_2 is found to be increased as size of the particle is reduced. So, using these expressions, it is possible to investigate particle size dependence of n, k, and other related constants like reflectivity.

References

1. A. Safaei, M.A. Shandiz, S. Sanjabiand, Z.H. Barber, J. Phys. Condens. Matter **19**, 216216 (2007)
2. K.K. Nanda, S.N. Sahu, S.N. Behera, Phys. Rev. A **66**, 013208 (2002)

3. H. Saka, Y. Nishikawa, T. Imura, Philos. Mag. A **57**, 895 (1988)
4. H.W. Sheng, K. Lu, E. Ma, Nanostruct. Mater. **10**, 865 (1998)
5. L. Grabaek et al., Phys. Rev. Lett. **64**, 934 (1990)
6. K. Chattopadhyay, R. Goswami, Prog. Mater Sci. **42**, 287 (1997)
7. J.M.D. Coey, *Rare-Earth Iron Permanent Magnets* (Oxford Science Publications, Clarendon Press, Oxford, 1996)
8. M.E. McHenry, M.A. Willard, D.E. Laughlin, Prog. Mater Sci. **44**, 291–433 (1999)
9. N.A. Frey, S. Peng, K. Cheng, S. Sun, ChemSoc Rev. **38**(9), 2532–2542 (2009)
10. S. Ayyappan, S. Mahadevan, P. Chandramohan, M. P. Srinivasan, J. Philip, B. Raj, J. Phys. Chem. C **114**, 6334–6341 (2010)
11. K. Mandal, S. Mitra, P.A. Kumar, Europhys. Lett. **75**(4), 618–623 (2006)
12. J.F. Hochepied, M.P. Pileni, J. Appl. Phys. **87**(5), 2472 (2000)
13. M.Y. Ge, L.Y. Han, U. Wiedwald, X.B. Xu, C. Wang, K. Kuepper, P. Ziemann, J.Z. Jiang, Nanotechnology **21**, 425702 (2010)
14. V. Blanco-Gutiérrez, M.J. Torralvo-Fernández, R. Sáez-Puche, J. Phys. Chem. C, **114**, 1789–1795 (2010)
15. H.D. Pfannes, Hyperfine Interact. **110**, 127–134 (1997)
16. P.G. Sanders, J.A. Eastman, J.R. Weertman, Acta Mater. **45**, 4019 (1997)
17. J.B. Wachtman (ed.), *Mechanical and Thermal Properties of Ceramics* (NBS Special Publication, NBS Washington, 1963), p. 139
18. J.K. MacKenzie, Proc. Phys. Soc. B **63**, 2 (1950)
19. H.S. Mansur, Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol. **2**, 113–129 (2010)
20. M.A. Meyers et al., Prog. Mater. Sci. **51**, 427–556 (2006)

Advanced Characterization Techniques

Chandan Kumar Ghosh

During past few decades, nanoparticles have attracted large number research scientist due to their exotic properties originating from their smaller size (less than 100 nm). Various methods and growth mechanism have been proposed to for nanoparticles having different morphologies like cubes, polyhedrons, rods, rings etc. Due to larger surface area and quantum effect, nanoparticles often possesses properties completely different from bulk those make it possible to use them in the field of single electron devices, nanoelectronics, drug delivery etc. They can be used to generate various predetermined structure in order to achieve some specific purpose. In this chapter, we'll first introduce few methods to synthesis nanoparticles and then some characterization techniques will be discussed in order to understand their physical properties.

1 Synthesis of Nanomaterials

In this section of the chapter, we have discussed different methods to synthesize nanostructure and to generate nanopattern. Synthesis procedures of the nanostructures are divided broadly into two categories: "top–down" and "bottom–up" methods. First method starts from large objects and reduces their dimension at nanoscale. The second starting from atomic or molecular levels produces nanoparticles. In some case, these nanoparticles are assembled into larger particles in order to obtain desirable applications. Depending on the principle of operation, synthesis methods are broadly classified into physical and chemical methods. Here, we will give outline of some physical methods like ball milling, evaporation, sputtering, chemical vapor deposition, etc., and few chemical methods such as sol–gel, chemical bath deposition, hydrothermal, sonochemical, etc. Subsequently, we have also discussed some lithographic techniques and those are being widely used to generate nanostructure pattern in the recent time.

C.K. Ghosh (✉)
School of Materials Science and Nanotechnology, Jadavpur University,
Kolkata 700032, West Bengal, India
e-mail: chandu_ju@yahoo.co.in

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_6

1.1 Physical Methods

1.1.1 Ball Milling

Whenever we are thinking of nanomaterials, we have to go through some specific and efficient synthesis processes. Ball milling or mechanical alloying is one of the low-cost process, used vividly in the field of powdered metallurgy, by means of which, we can generate very fine powder. The main mechanism involved here is the mechanical grinding. It consists of a hollow cylindrical shell, rotating about its own axis. The axis may be horizontal or at small angle to the horizontal. The hollow inner space is partially filled with steel or rubber balls, and the cylindrical inner surface has lining made of abrasion resistance coating (such as manganese steel). The material to be grinded is placed inside a hollow chamber. During rotation of the cylinder, balls create high energetic impacts, by means of which, we can generate fine powder from bulk system (top–down approach). But for nanoparticle synthesis, we need more energy and we have to opt for high-energy ball milling [1, 2]. During the high-energy ball milling process, the powder particles are subjected to high energetic impact. Sometimes, the energy is sufficient to initiate chemical reactions. The force of the impact plastically deforms the powder particles. The particles get the work hardened and fracture. The newly generated surfaces enable the particles

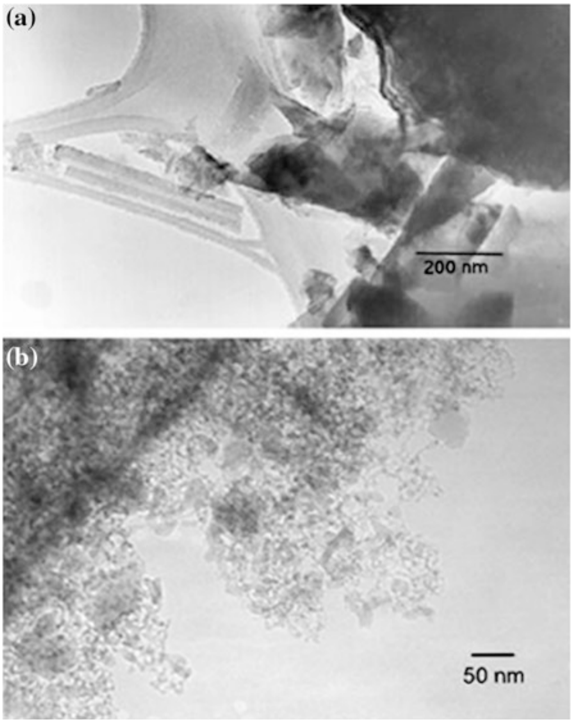
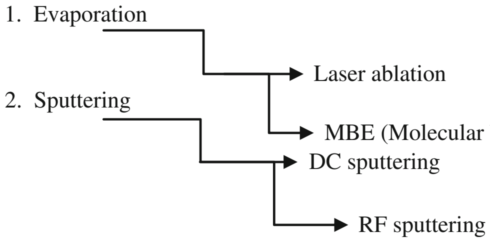


Fig. 1 Transmission electron micrographs of **a** unmilled crystalline α-AlF_3 **b** milled α-AlF_3 (reprinted with permission from Scholz et al. [3]). © ACS 2008

to weld together. Recently, few advanced milling technologies like SPEX shaker mills, Attritor mills, planetary ball milling, etc., techniques have been developed. Though this technique is very simple and cost-effective, it has several limitations. The main disadvantage of it is that we cannot generate non-agglomerated nanoparticle with homogeneous size distribution, or cannot control shape of the nanoparticles (Fig. 1).

1.1.2 Physical Vapor Deposition

Physical vapor deposition (PVD) is a typical technique for thin film formation procedure, where a target is used to transfer the source material to grow on the surface of a substrate. From the nomenclature we can easily understand that the process is dominated by physical methods, i.e., no chemical reaction takes place during the process. Depending on the procedure, we may broadly divide PVD as,



Evaporation

Evaporation is an age old procedure for thin film preparation. We may call it the father of thin film synthesis procedure. The most important aspect lies in the simplicity of the process. Basic principle is, we have to vaporize the source by thermal energy in a low pressure vacuum chamber and grow the material on the substrate which is connected with a holder (Fig. 2).

The key controlling parameter of this technique is the vapor pressure that is being generated simply by heating the source and at equilibrium, the expression for vapor pressure (P_e) can be written as (Ref: Guozhong Cao)

$$\ln P_e = -\frac{\Delta H_e}{R_g T} + C;$$

ΔH_e = Molar heat of evaporation, R_g = Gas constant, T = Temperature, C = Constant.

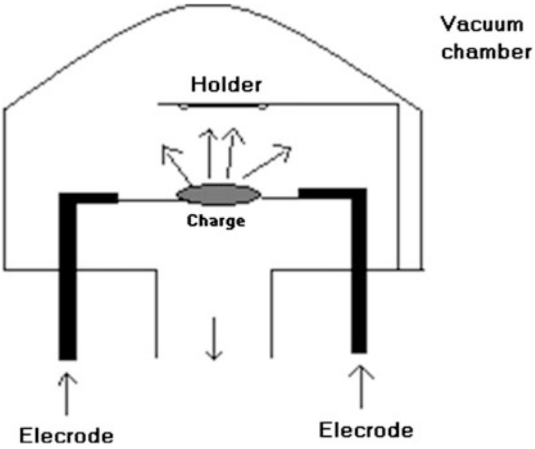


Fig. 2 Schematic diagram of evaporation system

Another obvious important parameter, involved in this process, is the evaporation rate (ϕ_e) which is governed by the equation [4]

$$\phi_e = \frac{\alpha_e N_a (P_e - P_h)}{2\pi m R_g T}$$

where α_e and N_a are coefficient rate and Avogadro number, respectively. P_h and m represent hydrostatic pressure and molecular weight, respectively. From the equation it is clear that the evaporation process depends on the molar composition. So in this process difficulties may arise when we go for film of the multicomponent with different molecular weights. Initially, very low pressure (10^{-3}–10^{-10} Torr) has to be created within the system to avoid the collision between atoms or molecules in the path between target and source. The transport phenomenon occurs along a straight forward way as the mean free path is very large compared to the target to substrate distance. The main limitation of the process is the uniform growth of the film over large area. Several modifications like the use of multiple sources, rotation of the substrate, loading the source, and substrate on the surface of a sphere have incorporated into this technique. In the recent time two techniques, namely, laser ablation and molecular beam epitaxy have been developed to get better control over material synthesis (Fig. 3).

Laser ablation is a modified evaporation technique that relies on the heating process by laser [6]. Generally, pulsed beam laser is used for the procedure. Depending on the type of material to be deposited, the wavelength of the laser is modified. Low energy conversion is the main drawback of the process (Fig. 4).

Molecular beam epitaxy (MBE) is a modified evaporation technique for single crystal growth [8]. The operation principle is the same as evaporation but here ultrahigh vacuum condition is used. Molecular beams are generated by heating the solid materials placed in the source cells, which is referred to as Knudsen cells. Generally, resistive heating is used for this purpose. The cells are aligned radially with the substrate.

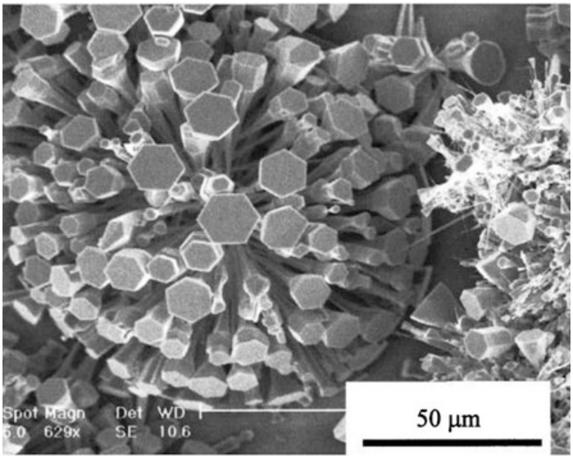


Fig. 3 SEM image of ZnO thin film by thermal evaporation (reprinted with permission from Yao et al. [5]). © AIP 2002

Though MBE is very expensive, it is popular because,

1. ultrahigh vacuum guarantees for negligible contamination.
2. any in situ characterization tools like X-ray photoelectric spectroscopy (XPS), auger electron spectroscopy (AES), reflection high-energy electron diffraction (RHEED), etc., can be attached to monitor and further characterization without altering the ambient condition.
3. use of single crystal substrate
4. individual evaporation rate control of different sources.
5. slow growth rate
6. low growth temperature
7. formation of hyperabrupt surface
8. precise control helps to grow 2D layers.

Sputtering

Sputtering is a special type of PVD process that is trusted by material scientists for a long period. The basic sputtering unit consists of a target or cathode, anode (substrate), and a sputtering medium where the target is bombarded by energetic particles. It only happens when the kinetic energy of the incoming particles is much higher than conventional thermal energies ($\gg$1 eV). It is commonly utilized for thin film deposition, etching, and analytical purposes. The easiest method to generate the bombardment is to ignite a glow discharge with target as cathode and substrate serving as anode. The attractive property of sputtering is its universal applicability. Since the coating material is transferred into the gas phase by impulse transfer and not by thermal excitation, therefor any material can be sputtered. Generally, argon is used as the sputtering gas. Depending on the source, two different sputtering techniques exist: (1) DC sputtering and (2) Radio frequency (RF) sputtering [9, 10].

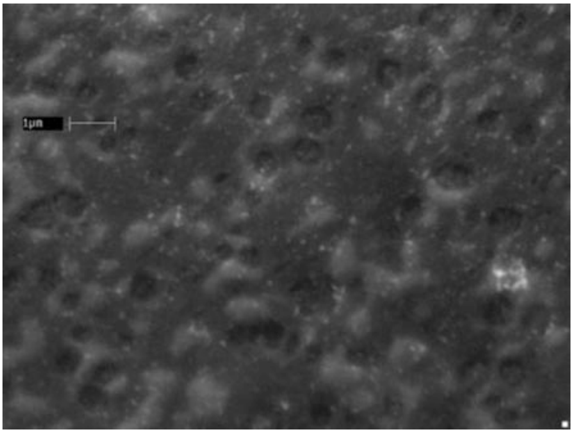


Fig. 4 ZnO Thin film prepared by laser ablation (*Courtesy* Thompson [7])

Though the methods are different, the underlying philosophy is almost same. DC sputtering system is the oldest sputtering technique. It contains a cathode (target), anode (substrate), and argon as sputtering medium. The pressure of the Ar is typically lying in the range 100 m Torr that causes initiation and discharge. In this technique, pressure of Ar is a very important factor that must not be too low because the gas phase reaction is very important, otherwise electron will directly hit the cathode. But it does not allow us to increase the pressure to a certain level that the number of collision becomes too high so that electrons might not get the desired energy. In the presence of an applied DC electric field (in K Vs/cm) between the electrodes, glow discharge initiates. The free electron resulting from the discharge is accelerated by the field, which simply knocked of some electrons from the neutral Ar atoms, creating Ar^+ ions. Now, these Ar^+ ions are attracted toward the cathode and hit it with its high momentum. So this is the process where momentum transformation happens. As a result, clusters from target get deposited on the substrate. In this DC sputtering method, the main criteria for target material that acts as a cathode is that it should be conducting in nature. For insulating materials, RF sputtering technique is adopted. Though typical RF varies between 5 and 30 MHz, but 13.56 MHz is reserved for this particular process by Federal Communications Commission (FCC). Now, if the cathode has low RF impedance, the target is self-biased to a negative potential and acts like DC source. The reason is very simple. Electrons have much higher mobility than the Ar. In such a high frequency (13.56 MHz) electrons will not face any problem to follow the periodic change in the potential (Fig. 5).

1.1.3 Chemical Vapor Deposition

As the name suggests, chemical vapor deposition (CVD) mainly involves chemical reactions that make it different from PVD processes [11]. In CVD, the reaction occurs in gas phase giving a nonvolatile solid, which can be deposited on a suitable

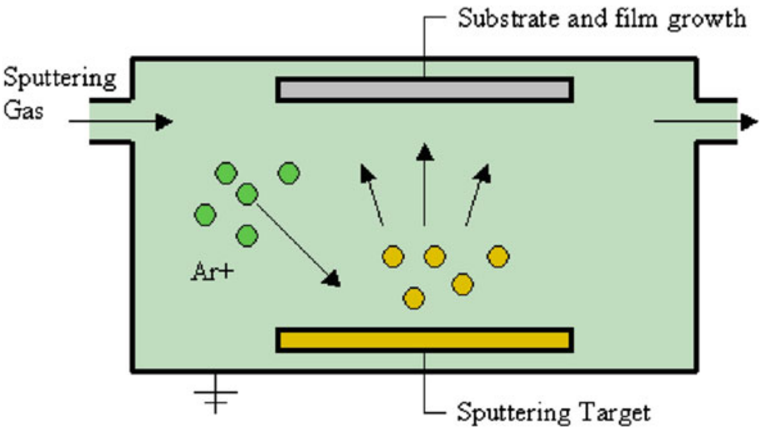


Fig. 5 Schematic diagram of sputtering (*Source* Wikipedia)

substrate to make it a thin film. The CVD includes wide varieties of reactions like pyrolysis, reduction, oxidation, disproportionation, reversible transfer, etc., depending on the activation of the CVD process, it is divided into several methods like:

- Thermal activation
  In thermally activated CVD, high temperature (>900 °C) is required. But in the case of organometallic components, this operates at relatively low temperature, called metal-organic CVD (MOCVD).
- Plasma activation
  Basically, it is low-temperature CVD, lies in the range of 300–500 °C. It is typically called as PECVD (plasma-enhanced CVD), where the plasma are excited by RF (100 kHz–40 MHz) or by microwave.
- Photon activation
  Photon-activated CVD involves the direct activation of reactant in the presence of shortwave ultraviolet radiation.

CVD is a very flexible technique that can be used to synthesize particular solid from different precursors and by varying the concentration of them. Also by varying the precursor composition and temperature from the same precursors we can have different films. Another interesting characteristic of CVD is that various morphologies can be generated using this method, unlike PVD, CVD does not require ultrahigh vacuum chambers. Though CVD is a nonequilibrium process, we have to go for equilibrium analysis to understand its intricacy. Actually, the reaction is governed by first and second law of thermodynamics that predict the feasibility of the reactions, its kinetics and controls the energies associated with it. The geometry of the reactor along with the chamber pressure also plays dominant roles. For most CVD system the pressure remains at 0.01 atmospheres, but in a special case, called low-pressure CVD, pressure lies in the range of $\sim$0.5–1 Torr. There are many types of CVD reactors. We can broadly divide them into hot wall and cold wall. Hot wall reactors are tubular in shape and resistive heating has been used all around. But in the case of cold wall reactors, inductive heating has been given to the substrate by graphite susceptors but wall remains cold with the help of air/water flow.

1.1.4 Atomic Layer Deposition (ALD)

Atomic layer deposition (ALD) is a newly introduced technology. It is a modified CVD process by which we can go for pristine control over the growth mechanism. It has a self-limiting growth nature, which cannot be done in any other CVD processes. That means we can create mono or few layer (2D/nearly 2D) films with the help of ALD. ALD mechanism is a combination of vapor phase self-assembly and surface reaction where the surface is activated by chemical reaction [12]. The precursor molecules are introduced in the chamber where they react with the activated surface and form chemical bond with the substrate. The precursor molecules do not react with each other so that a monolayer forms. If we further activate the newly formed monolayer, it will act as base and we may have more than one layer. Depending on the precursor, complex films can be easily generated by this technique.

1.2 Chemical Methods

1.2.1 Sol–Gel

The sol–gel technique is one of the attractive and versatile methods to synthesize a wide range of materials at extremely low cost [13]. It is also used to synthesize ceramic coating from solutions. In the case of synthesis of nanoparticles using this technique, surface area, pore volume, grain size can easily be tuned by careful control of chemistry of the sol–gel process. This technique was first identified by Ebelmen during hydrolysis and polycondensation of tetraethyl orthosilicate (TEOS). A ‘sol’ is defined as a stable dispersion of colloidal particles or polymers in a solvent, whereas colloid is defined as suspension of dispersed solid particles ($\sim$1–1000 nm) that does not involve agglomeration or sedimentation. On the other hand, gel consists of a three-dimensional continuous network, which encloses liquid phase. Sol–gel processes are categorized into aqueous based that involved water during reaction and alcohol based that excludes water. On the other hand, non-hydrolytic sol–gel process exists in nature that does not require any solvent. Sol–gel processes sometimes are also categorized into alkoxides and non-alkoxide groups according to volatility of the precursor material. Metal alkoxide is one of the best precursor materials and those are being widely used, because most of the metals are capable of forming metal alkoxides with general chemical formula $M(\mathrm{RO})_x$, M and R represent the metal and alkyl group, x is the valence state of the metal.

In order to obtain homogeneous coating or monodispersed nanoparticle using sol–gel process, homogeneous gel without precipitation is required. Generally, precipitation arises from physical agglomeration or chemical reaction of alkoxide with water or chelating agents that give insoluble hydroxides or organic salts. Sometimes, additives are used to increase the time of gelation to increase the stability of sol–gel derived product. The basic sol–gel reaction involves dispersed colloidal particles having very small size such that gravitational force on them is negligible and exhibits random Brownian motion within the fluid matrix. The interaction among them is dominated by short-range forces like Van der Walls attraction, surface charges, and extends only for a few nanometers. In order to counter these attractive forces, some repulsive forces are always present there. Depending on the way of formation of network structure, gel is classified as colloidal and polymeric gels. Gel forming from colloidal particles upon aggregation and condensation are linked into a three-dimensional network surrounded by liquid phase are termed as colloidal gels. Theses gels are formed from sol particles and are also known as corpuscular gel. Polymeric gels are formed from sub-colloidal chemical units like branched macromolecules where individual particles cannot be distinguished. It is defined as polymeric gel because their structures are generated through repetition of one or a few elementary units. The solubility of the macromolecules into water or alcohol or acid medium restricts the transformation of their chains into individual particles. The gel corresponding to subunits of colloidal chemical directly forms from the precursor solution without the intermediate occurrence of individual particles. The interactions within gel, in most of cases, are covalent in nature and the gelation process is found to be reversible. This sol–gel process depends on the precursor material, solvent, and catalysts (if necessary). The precursor, consisting of metal or metalloid elements, surrounded by various ligands (if necessary), generally guides the formation of colloidal particle. Alkoxides are mostly used for this purpose. Hydrolysis and polycondensation are the two important steps, involved in the sol–gel process, those prefer water as solvent medium. To fasten the reaction kinetics catalysts that promote the hydrolysis of the precursors are used in some cases. The basic mechanism, involved in this sol–gel method to synthesize solid materials at liquid and low temperature, is polymerization process that establishes metal–OH–metal or metal–O–metal bridges between metallic atoms of the precursor materials. As a result, colloidal particles are formed. The next step involves the linking of these colloidal particles to build a three-dimensional network structure within solvent. The step of gelation arises when linking process enhances to an extent such that a giant spanning cluster is formed. At this point, the gel possesses high viscosity with negligible elasticity. As time goes on, further linking processes and chemical inclusion of the particles into the spanning cluster increases, leading to an enhancement of the elasticity of the sample. It has to be mentioned that acidic medium favors hydrolysis and condensation determines the sol–gel kinetics, whereas hydrolysis determines the kinetics in basic medium. Another important factor in this sol–gel reaction is the aging and it plays a crucial role when there still exist a large number of sol particles to be reacted. Sometimes, it is required to remove liquid from gel by drying that influences structures of the network. For example, if the gel is dried by evaporation, then the capillary forces come into play that shrinkages the gel and a new type of gel, called aerogel, is formed. On the other hand, if it is dried under supercritical conditions, the network structure with high porosity evolves. Though the sol–gel process is very simple and cost-effective for synthesis of nanoparticles, but the major disadvantage of it is that it is very much substrate dependent in case of thin film synthesis.

1.2.2 Chemical Bath Deposition

Chemical bath deposition (CBD), also called chemical solution method, is one of the solution phase methods useful for the preparation of low-cost large area thin film from aqueous solutions (in general) [14]. It is widely used for the deposition of various metal chalcogenide thin films. It produces good deposits on suitable substrates by the controlled precipitation of the compounds from the solution. It does not require high voltage equipment, works at room temperature, and hence it is inexpensive. The only requirement for this deposition route is an aqueous solution consisting of a few common chemicals and a substrate for the film to be deposited. It often suffers from a lack of reproducibility in comparison with other chemical processes. However, by proper and careful optimization of the growth parameters, one can achieve reasonable reproducibility. The major problem of the CBD method is the inefficiency of the process which converts the precursor materials into useful deposits. Film formation may take place from two distinct mechanisms. The first one is a growth mechanism involving the reaction of atomic species at the surface. It corresponds to an atom-by-atom process, also called "ion-by-ion." The second deposition mechanism is associated to the agglomeration of colloids formed in the solution that may be considered a cluster-by-cluster growth process. In practical situations, both processes may interact, leading to films where colloids are included in the growing film. In other words, the formation of nuclei can occur by heterogeneous nucleation on the substrate or by homogeneous nucleation in the bulk solution, respectively. Homogeneous nucleation leads to rapid formation of large particles throughout the solution as precipitate. Conversely, heterogeneous nucleation occurs at the substrate surface and particles grow slowly to form a film. The heterogeneous growth of the film on the substrate is disrupted by the competing homogeneous reaction in solution and the subsequent deposition on the surface of the vessel containing the solution bath. Though CBD has many advantages, it suffers from some few drawbacks. In case of a CBD process, the heat needed for chemical reaction is supplied from the solution bath to the sample surface, resulting in both heterogeneous nucleations at the surface as well as homogeneous particle formation in the bath. The growth solution changes as a function of time that results in difficulty to control the thickness. The depletion of reactants also limits the achievable thickness. Moreover, the unequal bath-to-surface volume used to form the desired thin film generates a lot of waste and creates defects in devices.

1.2.3 Sonochemical and Arrested Precipitation Technique

Sonochemistry relies on the synthesis of nanoparticles using ultrasound. The main mechanism involved in this method is the phenomenon of acoustic cavitation effects in liquids. Cavitation bubbles are nothing but vacuum bubbles those are produced by periodic compression and re-refraction of sound wave propagating within the liquid. Ultrasonic transducer is mainly used to generate sound wave. Vacuum bubbles take part in the formation of nanoparticles.

Arrested precipitation is one of the controlled technique by which nearly monodisperse nanoparticle can be prepared. Initially, seeds of very small particles are prepared that act as seed particle to form bigger particle. It has been well studied from thermodynamics that smaller particles possess less stability than bigger particles. So, smaller crystallites have a trend to dissolve into their respective ions and try to recrystallize on comparatively stable larger crystallites. The process, involved in this growing process, is the Ostwald ripening. To avoid secondary growth, solvents like acetonitrile, etc., are used as capping agents. As an example, tri-n-octylphosphine selenide, known as TOPSe, is used as capping agent to synthesis CdSe nanocrystalline particles [15] (Fig. 6).

In most of the time, chemical reactions occur so fast that they cannot be controlled easily. To reduce such rate, concentration of the precursor, known as the 'nucleation threshold,' is lowered very much.

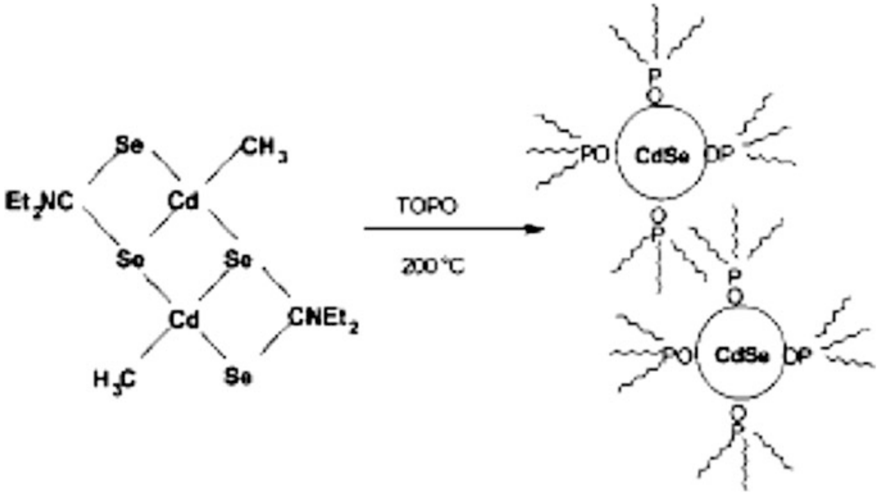


Fig. 6 Formation of CdSe nanoparticles

1.2.4 Photochemical Method

This method of synthesis occurs in the presence of light, generally UV or visible light. The chemical reaction occurs in the presence of photons. Photochemical reactions generally yield reduced nanoparticles of metals like Ag, Au, etc. For a typical example of synthesis of gold nanoparticles, aqueous solution of chloroauric acid gives Au^+ ions and those are reduced in the presence of UV light to form metallic Au nanoparticles. The reduction reaction is as follows:

$$Au^+ \rightarrow Au^0 + 1e^-$$

The color of the sample that originates from surface plasmon depends on the size of the particles so formed.

1.2.5 Hydrothermal

The term "hydrothermal" is difficult to define, based on the Greek word, "hydrous" means water and "thermal" means heat. One of the accepted statements for hydrothermal defines it as any heterogeneous chemical reaction that occurs in the presence of a solvent media at above the room temperature (>25 °C) and pressure levels greater than 0.1 MPa in a closed system, at these conditions does not matter whether the solvent is aqueous (water) or nonaqueous (alcohol). Water is one of the most important solvent present in nature in abundant amount and has remarkable properties as a reaction medium under hydrothermal conditions [16]. The crystallization process of solid phases under hydrothermal conditions is usually conducted at auto-generous pressure, achieving a particular saturated vapor pressure of the solution at the specific temperature and composition of the hydrothermal solution (Fig. 7).

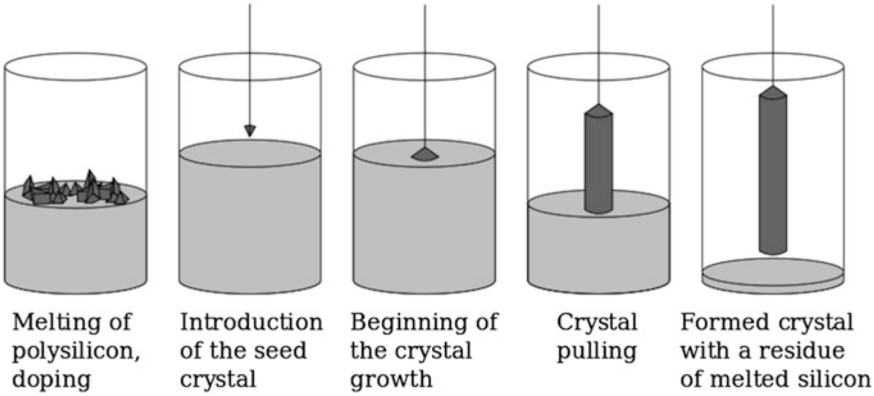

Melting of polysilicon, doping
Introduction of the seed crystal
Beginning of the crystal growth
Crystal pulling
Formed crystal with a residue of melted silicon

Fig. 7 Schematic diagram of Czochralski crystal growth process (*Source* Wikipedia)

In this synthesis techniques, normally temperature <350 °C and pressures >50 MPa are maintained. Crystal growth under hydrothermal conditions requires a reaction vessel called an autoclave. An ideal hydrothermal autoclave should have the following characteristics:

1. it should inert to acids, bases, and oxidizing agents.
2. it should be easily assemble and dissemble.
3. it should have sufficient length to obtain a desired temperature gradient.
4. it should be leak-proof at desired temperature and pressure.
5. it should bear high pressure and temperature for long duration of time.

Hydrothermal synthesis offers many advantages over other conventional and nonconventional synthesis methods. Unlike many advanced methods that can prepare a large variety of forms, the respective costs for instrumentation, energy, and precursors are less for hydrothermal methods. The low reaction temperatures also avoid other problems encountered with high temperature processes (Czochralski method, Bridgeman method) such as poor stoichiometric control due to volatilization of components and stress-induced defects (e.g., microcracks) caused by phase transformations that occur as the phosphor is cooled to room temperature. Moreover, the ability to precipitate the powders directly from solution regulates the rate and uniformity of nucleation, growth and aging, which affects size, morphology, and aggregation control that is not possible with many synthesis processes. A major advantage of hydrothermal synthesis is that this method can be hybridized with other processes like microwave, electrochemistry, ultrasound, mechanochemistry, optical radiation, and hot pressing to gain advantages such as enhancement of reaction kinetics and increase ability to make new materials. A great amount of work has been done to enhance hydrothermal synthesis by hybridizing this method with many other processes. This facile method does not need any seed, catalyst, harmful and expensive surfactant or template thus it is promising for large scale and low-cost production with high-quality crystals.

1.3 Lithography

The word of ‘lithography’ comes from two Greek words, “lithos” (means ‘stone’) and “graphein” (means ‘writing a pattern on stone’). In perspective of recent technology for device fabrication, lithography is one of the suitable methods. The main principle involved in lithography is that a resist material is spun onto the wafer containing a thin layer on it. Then resist is selectively exposed to different external excitation through a patterned mask that contains information about the particulars being fabricated. The resist is finally developed. So we can say in short that the lithography is the technique of transferring geometric designed patterns from a mask onto a wafer (mainly Si is used for this purpose). According to Moore’s law, the demand for the fabrication of micro- and nanoelectronic devices is increasing rapidly during last two decades and lithography is found to fulfill the demands with its essentiality in the production of integrated micro- and nanoelectronic devices. The main driving force of lithographic advancement is also found to be the reduced cost. This technique was developed at the early stage of time from the shadow printing that was a well-known process in the printing industry and also for patterning printed circuit board. Rubylith was the master pattern substrate, used in 1960, that consists of a layer of heavy transparent acetate, onto which ruby red layer lightly adheres. The second layer is cut to generate the pattern. In recent time, several forms of lithographic techniques like photolithography, X-ray lithography, electron beam lithography (EBL), ion beam lithography etc. have been developed for ongoing inexorable trend in electronic device development. Photolithography also called the optical lithography was the first lithographic technique that has been widely used during last three decades to fabricate devices with length scale between 15 μm and 180 nm. This method mainly relies on the formation of images in the presence of radiation in a photo resist. This lithographic technique has been developed on the basis of etching resist in the presence of light, either directly or with projected image having optical mask. In recent time, with advancement of technology, ion beam lithography, X-ray lithography, EBL etc. have been emerged with better control and high resolution.

1.3.1 X-ray Lithography

X-ray lithography is one of the modern lithographic techniques that is being widely used in recent time to transfers the geometric patterns from a mask to the surface of the wafer (mainly Si). International Business Machines (IBM) first developed X-ray lithographic technique in the year 1969 and generated metal structure ~20 μm. In this lithographic technique, X-ray from synchrotron radiation is used to generate the pattern [17]. Similar to photolithography, the first step, involved in this lithographic technique, is the exposure of thick layer of resist via patterned mask to highly intensed X-ray beam. The pattern is then exposed to X-ray such that it gets etching into the resist substrate that results negative replica of the mask pattern. In order to remove unwanted material and resist materials (in case of free standing material), chemical treatment is carried out. The mask materials, generally used in X-ray lithography, are made of materials with low atomic mass (polymethyl methacrylate (PMMA) is used as resist material). In this context, it has to be mentioned that the mask materials with high atomic number can't be used due to attenuation of X-ray. Primarily the requirement of choosing substrate is that it should be conducting in nature. Resolution of X-ray lithography is increased by decreasing wavelength of the X-ray. But the major problem that is inherent in this lithographic technique is the proper alignment of the mask with resist, since no visible light can penetrate through X-ray membrane.

1.3.2 Electron Beam Lithography

Electron beam lithography (EBL), after derivation from SEM, is routinely used in order to write directly on resist since early 1960. Direct writing scheme, employed here, uses a heavily focused Gaussian shaped electron beam that moves on the wafer upon which it produces one pixel at a time. The movement of the electron beam can be either raster type or vector type. This lithographic system consists of a source of electrons, focusing optics for electrons, blanker (to either turn off or on the electron beam), generator to move the electron beam and a mechanical system to hold the substrate. It is now possible using direct writing method to produce electronic devices $\sim$10 nm with suitable combination of proper etching and fabrication without any projection optics or mask. In recent time, significant efforts have been given to develop electron beam having different shape to enhance the exposure speed and to increase the application opportunity by integrating with other nanoscale processes. The important technique for direct writing, involved in this lithographic process, is the lift-off process. It is an additive process and involves few steps: exposure of the resist by electron beam, development of the exposed resist and removal of the solvent. Finally the lift-off process is accompanied by soaking the substrate to remove remaining resist and unreacted materials. This process is frequently used in most semiconductors device fabrication due to its preferences of thick layer for forming its undercut profile and adhesion with substrate [18].

The other important scheme, involved in EBL, is the projection printing that was started from Bell laboratories in the year 1989. The basic principle, involved in this EBL process, is he scattering with angular limitation in projection EBL (known as SCALPEL). The mask, used here, is the membrane with low atomic number upon which a layer of pattern with high atomic number is deposited. When electrons incidents on such mask, they are scattered at higher angle by atomic layer with high atomic number compare to other. An aperture, kept at the focal plane of the projection lens, allows only those electrons, scattered at lower angles while the other electrons are blocked. The unblocked electrons form a contrast image on the wafer or substrate.

1.3.3 Ion Beam Lithography

Compare to other lithographic techniques, ion beam lithography (IBL), where ions are used to scan, is less known, but possesses mush versatility and capability. Working principle of IBL is similar to that of EBL, but shows higher resolution ($\sim$2 nm) compare to that of EBL [19]. Much smaller de Broglie wavelength of ion compare to that of electron increases the resolution of IBL. In contrast to EBL, IBL is free from back scattering of electron, hence much higher resolution is obtained. Pixel size of IBL is roughly equal to that of beam spot size having no exposure between them. As a consequence, IBL shows much smaller dwell time on each pixel. The main advantage of IBL is that it can be used for direct writing on the resistless structure. Not only writing, ion implantation can also be used to fabricate locally doped mask layers that can generate pattern in any selective etching process. Ion beams, in some cases, are also used to initiate the chemical reaction. The main mechanism, involved in IBL, is the ion induced modification of the resist depending on the interaction between ion and resist. Normally, hydrogen or helium gas is used in IBL.

2 Characterization of Nanomaterials

Not only syntheses, in order to better understand the property of the nanoparticles, they have to be characterized by different techniques. In early stage of time, peoples investigated materials property with visual inspection. For example, color was used to check mechanical property of quenched steel. With advancement of science and technology, new techniques have been developed to study materials property. As an example, optical microscope with lens system is used to study microstructure of materials science. Later more powerful techniques with higher resolution compare to visible light have been developed to investigate nanomaterials. In this subsection, we'll introduce some basic characterization techniques such as X-ray diffraction, scanning electron microscope (SEM), transmission electron microscope, scanning tunneling microscope, atomic force microscope etc. of nanoparticles. X-ray diffraction is used to investigate crystallinity of the nanomaterials and to identify the phase purity of the materials. We'll give some outline of different microscopic techniques those are widely used to study the microstructure, morphology of the synthesized nanoparticles. Basic principles of some optical characterization methods like UV–V is spectroscopy, ellipsometer will also be discussed.

2.1 Characterization by X-ray Diffraction

We know from basic knowledge of optics that diffraction occurs when electromagnetic radiations are incident on any periodic arrangement of obstacles having its periodicity comparable to the wavelength of the incident radiation. Crystalline materials are characterized by periodic arrangements of atoms. We may consider those periodic arrangements as plane through them. Since wavelength of X-ray (characteristic X-ray is used) is comparable to that of interplanar spacing of crystalline material, so diffraction effects can be observed. This diffraction phenomenon is generally used to determine fine structure of materials. This phenomenon is also used for chemical analysis, stress determination, particle size calculation etc. Though the phenomenon is diffraction but it can be understood from Bragg's law that corresponds to reflection of X-ray from successive planes of crystalline materials. It is one of the fundamental laws in X-ray diffraction that provides the

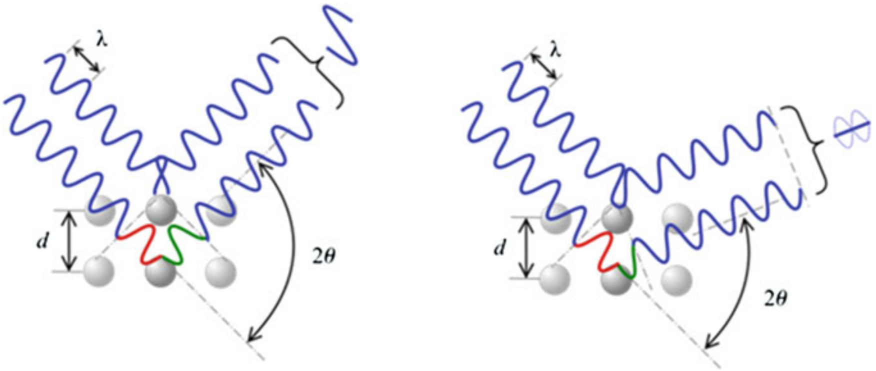


Fig. 8 Constructive and destructive interference of reflected X-ray (*Source* http://en.wikipedia.org/wiki/Bragg's_law)

relation between the wavelength of the incident X-ray and the incident angle and determines the diffraction pattern. As shown in the schematic diagram, X-rays, reflected from two successive planes, possesses a path difference of $2d \sin \theta$, where d and θ represent interplanar spacing and angle of the radiation with sample surface. If this path difference is equal to an integral multiple of λ (wavelength of the incident X-ray), then they will interfere constructively resulting maximum intensity, otherwise their interference is destructive giving zero intensity. The condition of interference maxima i.e., $2d \sin \theta = n\lambda$ ('n' is any integer) is known as Bragg's law [20] (Fig. 8).

The basic principle, involved in this experimental technique, is the measurement of intensity as a function of θ. Experimentally, intensity is measured as function of 2θ i.e., angle between incident and reflected X-rays. One typical diffraction pattern is shown in Fig. 5.9. Presence of different peaks is attributed to the satisfaction of Bragg's law for different planes. Intensity of the peaks depends on the number of electrons within atom and on the type and arrangement of atoms within unit cell of the crystalline materials. It has to be mentioned that not all planes show diffraction pattern. Actually the plane that will correspond diffraction maxima is determined by atomic form factor and structure factor. One main objective of this technique is phase identification. It is well known that each crystalline material possesses characteristic interplanar spacings. Phase identification is carried out by comparing the experimentally measured interplanar spacing with their standard value. Instead of δ-like pattern of diffraction peak in ideal case, its experimental pattern possesses finite width. There are several factors such as finite size of particle, strain etc. responsible for such finite width. Thus we may use this finite width to determine these parameters. For these purposes, let us consider a diffraction peak at angle 2θ with intensity $I_{\max}$. Let $2\theta_1$ ($<2\theta$) and $2\boldsymbol{\theta}_2$ ($>2\theta$) be the two angles beyond which the diffraction pattern corresponds to zero intensity. Thus using Bragg's law for two extreme planes, separated by 't', we may write,

$$2t\sin\theta_1 = (m-1)\,\lambda \text{ and } 2t\sin\theta_2 = (m+1)\lambda,$$

where '*m*' represents any integer number.

Subtracting one from another, it may be written,

$$t(\sin\theta_2 - \sin\theta_1) = \lambda$$
$$\text{i.e., } 2t\cos\left(\frac{\theta_2+\theta_1}{2}\right)\sin\left(\frac{\theta_2-\theta_1}{2}\right) = \lambda$$

with $\frac{\theta_2+\theta_1}{2} = \theta$ and $\theta_2 - \theta_1 = \beta$ presenting full width at half maxima (FWHM) of the diffraction pattern, we may write the above expression as follows,

$$t\cos\theta\,\sin\beta = \lambda$$

If β is expressed in radian, the above expression may be written as,

$$t\cos\theta\,\beta = \lambda$$
$$\text{i.e., } t = \frac{\lambda}{\beta\cos\theta}$$

More rigorous calculation gives

$$t = \frac{0.9\lambda}{\beta\cos\theta}$$

This expression, after derivation by Scherrer, is well known as Scherrer's relation and is mostly used to calculate particle size from diffraction pattern. The above equation has been derived considering only finite size of the particle that causes broadening of FWHM. Variation of the lattice parameter may be another source that enhances FWHM. To incorporate this, we consider Bragg's law $2d\sin\theta = \lambda(n=1)$. After differentiating above equation with respect to θ, we may write,

$$\frac{\Delta d}{d} + \frac{\cos\theta}{\sin\theta}\Delta\theta = 0$$
$$\text{i.e., } \Delta\theta = -\frac{\sin\theta}{\cos\theta}\frac{\Delta d}{d}$$

$\frac{\Delta d}{d}$ is fractional variation in lattice parameter that may be defined as strain (ε), developed within the system. Considering only magnitude of the strain contribution, we may write total FWHM,

$$\beta = \frac{\lambda}{t\cos\theta} + \frac{\sin\theta}{\cos\theta}\varepsilon$$

$$\text{i.e., } \beta\cos\theta = \frac{\lambda}{t} + \varepsilon\sin\theta$$

From above equation, it is clear that $\beta\cos\theta$ versus $\sin\theta$ for different peaks must be a straight line and the plot is known as Williamson–Hall plot. From slope of the curve we may determine the strain and from intercept with $\beta\cos\theta$ we may extract particle size. Cu K_α (λ = 1.5404 Å) and scintillation counter are mostly used as a source of radiation and X-ray detector.

2.2 Characterization by Scanning Electron Microscope/Field Emission Scanning Electron Microscope

Scanning electron microscope (SEM)/field emission scanning electron microscope (FESEM) is a microscope to visualize very small topographic details of the sample surface [21]. Researchers in nanoscience, material science, biology, chemistry and physics employs this technique to observe structures that may be as small as 1 nm (=billion of a millimeter). Electrons, liberated from a field emission source, are allowed to pass through high vacuum column in the presence of a voltage gradient (typically ~30 kV). These beams of electrons are focused by electronic lenses, called electromagnetic lenses (combination of electric field and magnetic field), to produce a narrow scanning beam of electron having diameter 1–10 nm (called electron probe). In conventional SEM/FESEM, the electron probe is used to scan the specimen horizontally in two perpendicular directions. This is generally carried out by two sawtooth wave generators. Scanning rate along one direction is much greater than that of other. During scanning along one direction (say x direction), the electron probe moves in a straight line (known as line scan), after reaching the end point it is again deflected back along x-axis as quickly as possible. In this subsequent time, probe moves slightly along other direction (say y direction). Again the probe moves along x-direction. The entire procedure is known as raster scanning that includes the beam to cover a rectangular area on the specimen. When these high energetic electrons enter (called primary electron) into the solid, they suffer from both elastic and inelastic scattering. The energy lost by primary electron during inelastic scattering is gained by atomic electrons. If this energy is sufficient for them, they escape from atom and travel freely through the solid specimen (called secondary electron). Secondary electrons escape from the sample into vacuum if they are being generated very close to the sample surface within some depth, known as escape depth. Since escape depth is the characteristic property of the material, the secondary electrons carries the signature of the surface structure (topography) of the specimen, rather than underlying structure. In electron microscope, signal from secondary electrons are collected by the photomultiplier tube for imaging purpose. Output of the two probe generators and image are synchronized to a display device i.e., there exists one-to-one correspondence between probe generators and signal of secondary electrons. The contrast in the images arises from the variation of number of secondary electron, collected at photomultiplier tube. In this context, it has to be mentioned that the following conditions as given by James Clerk Maxwell must be satisfied between object and image:

(a) for each object there should be one image point.
(b) geometrically the object and image must be similar.
(c) for any planer object perpendicular to the optic axis, corresponding image must be planer and perpendicular to the optic axis.

Image magnification in electron microscope is defined as the ratio of scan distance in the image over scan distance on the specimen. In modern electron microscope, the scan signals are digitalized by computer controlled circuit. The waveforms along the two dimensions are actually given in the staircase like function having m and n levels respectively. This process divides the whole image into ‘mn’ picture element (called pixel) and electron probe spends certain time at each pixel before jumping to next pixel. The main advantage of such digitalized scanning system is that the imaging computer can record the intensity information at each pixel point. The other advantage of this process is that the image size is kept fixed whereas the magnification is increased by reducing the scan pixel. Just like optical microscopes, electron microscopes also suffer from several aberrations like spherical aberration, axial aberration, axial astigmatism etc. that reduce spatial resolution of the objects.

According to operational principle, there exist two types of electron microscope namely SEM and FESEM. In SEM, electrons are produced by thermionic process in which the electron emitter is made of with V-shaped tungsten wire that emits electron in the presence of high dc current. On the other hand, field emission process is involved in the FESEM. The emitter is also consists of a single-crystal tungsten wire fashioned into a sharp point and spot welded to a tungsten hairpin. The significance of the small tip radius, about 100 nm or less, is that an electric field can be concentrated to an extreme level. If the tip is held at negative 3–5 kV relative to the anode, the applied electric field at the tip is so strong that the potential barrier for electrons becomes narrow in width. These narrow barriers allow electrons to “tunnel” directly through the barrier and leave the cathode without requiring any thermal energy to lift them over the work function barrier.

In order to take photograph by SEM, the primary requirement for sample is that it should be conducting in nature. For metallic or conducting sample, there exist no problem, but for insulators, in order get clear image, it should be coated by extremely thin layer (1.5–3.0 nm) of conducting material like gold or gold palladium. Biological material, however, requires a prefixation, e.g., with cold slush nitrogen (cryofixation) or with chemical compounds. This particular microscope consists of a special cryo-unit where frozen objects are fractured and coated for direct observation in the FESEM. Chemically fixed material needs first to be washed and dried below the critical point to avoid damage of the fine structures due to surface tension. Coating is then performed in a separate device (Fig. 9).

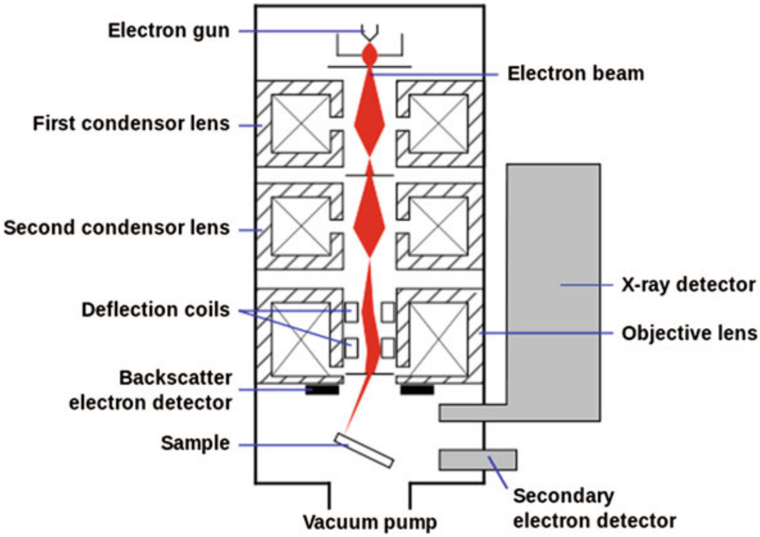


Fig. 9 Schematic diagram of the SEM apparatus (*Source* Wikipedia)

2.3 Energy Dispersive X-ray Spectroscopy

Energy Dispersive X-ray Spectroscopy is a basic technique for elemental analysis of materials. It is based on the fact that every element has a unique electronic structure and gives a distinct set of peaks in its X-ray emission spectrum. The basic idea behind its operation can be described as given hereunder.

The fundamental structure of an atom is quite well known and it is represented by electrons occupying specific orbits (electron shells) surrounding the nucleus of an atom. Each orbit or shell in that atom represents a discreet energy level. Prior to excitation the atom can be considered at rest (though every atom vibrates continuously about its mean position at its lattice point) and the electrons remain in the ground state (state corresponding to minimum energy). The analysis begins with the sample being bombarded by highly charged particles (proton, electron or X-ray). These highly charged particles on collision with an electron may ‘knock’ it out of its shell, resulting in the formation of an electron hole in its place. As a result, an electron from a higher energy shell will come into occupy its place and in the process releases energy in the form of X-ray. The energy of the emitted ray is equivalent to the difference of energy between the higher energy shell (from which the electron came in) and the lower energy shell (from which the electron was knocked out). This emitted energy is unique to the corresponding electron shells and subsequently the atomic structure of the element. Hence, the element can be identified from the value of the emitted energy and thus the constituent elements of the sample can be known along with the atomic percentages of each element.

2.4 Transmission Electron Microscope (TEM)

When the electronsare accelerated by high potential difference, their wavelength becomes comparable to the interplanar spacing of the crystalline materials. If such electrons interact with crystalline materials, they can penetrate and in fact, are scattered, as in the case of X-ray diffraction. Then it was realized that these scattered electrons can be used to extract information, mainly size and shape of the materials after transmission [22]. In the case of crystalline material having periodic arrangement of atoms they produce diffraction pattern, called transmission electron diffraction. In the early stage of time, gas discharge was used as the source of electron, but in recent time, electrons are emitted from tungsten wire having V-shape by field emission process. Emitted electrons are allowed to enter into a vacuum chamber, created by either terbomolecular or ion pump. In this vacuum chamber, electrons are accelerated by potential, generated by electronic oscillator and step-up transformer, to their final kinetic energy. Then the electrons pass through two or more condenser lenses that focus the electrons beam onto the specimen. Condenser lenses are nothing but external magnetic field. The accelerated focused electrons are allowed to fall on the sample that should be very thin such that electrons can penetrate the sample. The sample stage is designed in such a way that it holds the specimen stationary; otherwise any movement would reduce its resolution. There are two types of sample stage, namely side-entry and top-entry. One important requirement for sample stage is that it would not disturb the vacuum. This is done by inserting the sample through airlock. It is a small chamber in which the sample is kept for some time before insertion into the main vacuum chamber. After transmitting through the sample, electrons are collected onto imaging lens that gives an enlarged image of sample or the diffraction pattern. Imaging lens system consists of several components like objective lens, objective aperture, objective stigmator, selected area aperture, intermediate lens and projector lens. The signal that is being collected is the electronic signal; to convert it into visible form phosphor screen, consists of metal plate coated with fluoresce powder that emits visible upon electron bombardment, is used. In recent time, phosphor screen is replaced by charged-couple diode (CCD) based sensor consisting of millions of silicon photodiode. Two completely contrasting imaging fields, namely bright-image field and dark-image field are produced by the scattered electrons on the screen. The portion of the view (such as hole or portion beyond edge area) that contains no sample does not scatter any electron, hence the portion appears to be bright, and hence this portion is referred as bright-field image. Instead of unscattered electrons, scattered electrons can also be used to produce image on the screen, called dark-field image. Strongly scattering of electrons that depends on atomic number produces bright spot. Whatever may be the case, contrast in the image appears due to variation in the number of electrons collected on the screen. For amorphous materials, image arises from the composition and thickness of sample that varies between different regions and the contrast is described by the mass-thickness contrast. Scattering from polycrystalline materials consisting of small crystallites, known as grains, is also determined by the orientation of the atomic rows and columns relative to the direction of incident electrons. Thus alternatively we may conclude that using scattered electrons it is possible determine atomic arrangements thus interplanar spacing of crystalline materials (lattice fringe). In fact scattering of electrons from periodic arrangements of atoms produces diffraction pattern like X-ray diffraction. For a particular orientation, Bragg's law will be satisfied by crystallite planes and the material will strongly diffract the incident electrons. If this diffraction pattern (in form of either ring or spot on the screen) is collected at angles higher than that of the optical aperture, the crystallite will appear to be dark for planes satisfying Bragg's law and bright for planes not satisfying Bragg's law. Rings are basically constituted of large number of spots, each arising from reflection corresponding to the individual crystallites. Radial distribution of the spot that depends on the distance of the spots from center is determined by the interplanar spacing of the diffracting planes, whereas azimuthal distribution of the spots is dependent on the orientation of the planes with respect to direction of the incident beam. Thus from radial distribution of the spots, it is now possible to determine the interplanar spacing of the sample. It has been stated that the contrast in the TEM image arises from the variation of electron density in the case of polycrystalline material, but for single crystalline material, electron density may vary due to defects related to atomic vacancy, dislocation, stacking fault etc. So using TEM, it is also possible to investigate these defects of single crystalline materials.

2.5 Scanning Tunneling Microscopy

Scanning tunneling microscopy (STM) that was invented by Binning and Rohrer in the year 1981 has been proven to be a sophisticated and advanced characterization tools to investigate electronic structure of solid state nanostructure materials [23]. In recent time, it becomes also possible to image surface and local electronic structure associated with it. Basic principle inherent in this technique is the mapping of the surface by tunneling current between surface and STM. STM consists of sharp probing tip, one piezoelectric scanning system, positioning system and electronic control unit. Piezoelectric system is used to control both vertical and lateral movement of the scanning tip. The process is very much sensitive to vibration; therefore the entire measurement must be vibration free. The purpose of the coarse positioning system is to control the tunneling current. Electronic control system is generally used to measure the tunneling current and to control scanner with the help of a feedback circuit. The feedback circuit takes the output signal from the current pre-amp, and then it compares with the signal level in present value. The final step involves the response feedback voltage depending on the user-defined feedback parameters. Feedback voltage is sent through an analog to digital converter to the high voltage amplifier that magnifies the input signal to control the piezo scanner. Current flows between sample and tip electrode when the gap between them is of the order of several Å and a bias voltage is applied. Quantum mechanical calculation shows that the tunneling probability of the electrons depends on the separation between tip and sample surface and the corresponding tunneling current (I) α e^{-2kz}, where z is the thickness of the insulator and $k = \sqrt{2m(V-E)}/h$ where m = effective mass of electron, V = Electrostatic potential of the potential barrier (insulator), E is the energy of electron, h is the Planck's constant. It was observed experimentally that 0.1 nm variation in the gap results a change in current by an order of magnitude, hence STM possesses high tip sensitivity. The atomic resolution of the image along X–Y direction is 0.01 nm and 0.002 nm along Z direction. From differential conductance (dI/dV), the average DOS of the sample can be calculated. The tip is made up of metal or metal alloys like tungsten or platinum–iridium and it is mounted on a three dimensional array of piezoelectric material and it should be very sharp (Fig. 10).

Two different operational modes exists in this STM techniques:-

1. **Constant height mode**—In this mode the tip height above the surface is kept constant. So during scanning a profile of periodic variation of separation of tip and the surface atom is generated due to variation of tip position. The tunneling current will be large when the tip is just above one atom on the other hand the current will be low when the tip is above the gap between the two atoms. Thus this mode provides the direct image of the surface. The image is formed by plotting tunneling current versus tip position.
2. **Constant current mode**—In this mode the tunneling current is maintained constant so the image is formed due to the variation of the tip height. In this mode the tip is slightly move downward when it founds the gap and upward when it founds the atom. The image is formed by plotting the tip height versus the lateral tip position (Fig. 11).

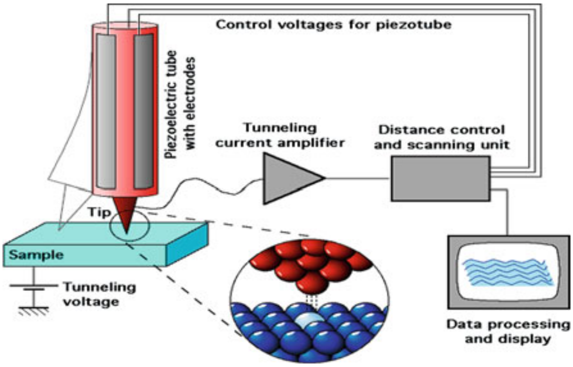

Control voltages for piezotube
Piezoelectric tube with electrodes
Tunneling current amplifier
Distance control and scanning unit
Tip
Sample
Tunneling voltage
Data processing and display

Fig. 10 Schematic of scanning tunneling microscope ("*Figure* Michael Schmid, TU Wien", under creative commons attribution ShareAlike 2.0 Austria license.)

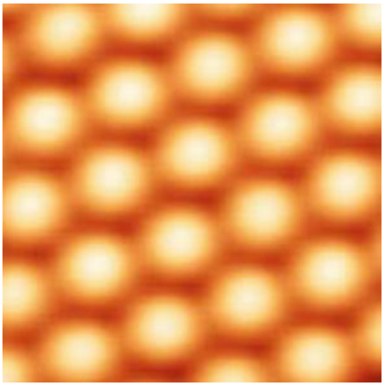


Fig. 11 The silicon atoms on the surface of a crystal of silicon carbide (SiC). Image obtained using an STM. (Guillaume Baffou—Image de microscopie à effet tunnel réalisée au lppm à Orsay under creative commons attribution ShareAlike3.0 license) (*Source* www.wikipedia.org)

2.6 Atomic Force Microscopy

The operation of STM is limited to conducting substrate. In order to scan non-conducting substrate to measure topography, atomic force microscopy (AFM) comes into picture [24]. In recent time, it is also utilized to measure the force. The basic AFM instrumentation consists of a cantilever with a very small and sharp tip attached to it, a laser diode that is used to measures the deflection of tip, a mirror (to deflect the laser light), a feedback circuit (provides the control of the position of tip by the regulation of the force), photodiodes (for indicating the position of the tip) and three dimensional positioning sample stage made of array of piezoelectric material. Lithography technique is employed to fabricate the AFM tip. The cantilever has small mass so it has high flexibility and exerts less downward forces on the sample that results less distortion and damage in the topography. During topographic height measurement, cantilever is pushed and pulled back. It gets deflection during the measurement that is proportional to the force between tip and the sample. There exist various methods like optical, piezo resistive methods, tunneling current measurement etc. to detect deflection of the cantilever. First developed cantilever consists of gold foil that contains diamond tip. Recently, Si, SiO_2, Si_3O_4 are used as cantilever. Since deflection of the AFM tip is proportional to the interacting force between tip and the sample (separation $\sim$0.1–100 nm), thus it can be used to know the interacting forces like van der Waals (vdW) force, electrostatic force, chemical force etc. In this context, force spectroscopy has been developed that gives information local interacting force having spatial resolution on the nanometer scale, acting between tip and sample. In contrast to horizontal scan as in case of topographic measurements, force spectroscopy relies on the vertical scan of the sample. These are the basic operational principle of static AFM mode. In contrast to static AFM mode, there exists another mode in the form of dynamic

AFM mode that relies on the change of either natural frequency of the tip or vibrating amplitude or phase of the cantilever when they are brought close to the sample surface (Figs. 12 and 13).

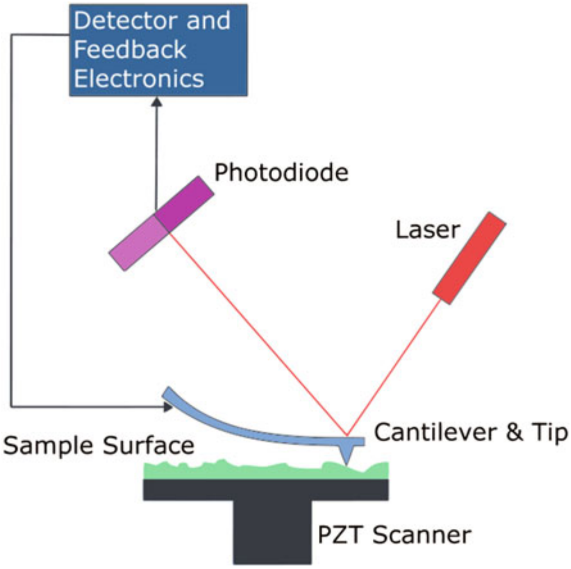


Fig. 12 Schematic of atomic force microscope (*Source* www.wikipedia.org)

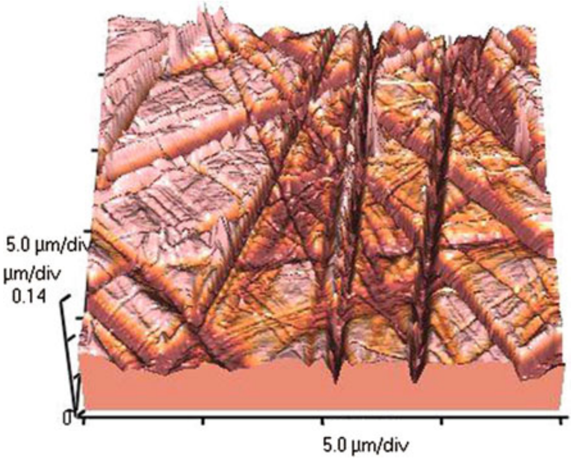


Fig. 13 Atomic force microscope topographical scan of a glass surface. The micro and nanoscale features of the glass can be observed, portraying the roughness of the material. The image space is (x, y, z) = (20 µm × 20 µm × 420 nm). The AFM used was the Veeco di CP-II, scanned in contact mode. Constructed at the Nanorobotics Laboratory at Carnegie Mellon University (http://nanolab.me.cmu.edu) (*Source* www.wikipedia.org)

2.7 Optical Characterization by UV-Visible Spectroscopy

Generally band gap is determined by UV-Visible spectrophotometer that consists of a light source (deuterium lamp for UV light and halogen lamp for visible), grating, rotating discs, slit, mirrors, sample and reference cells and detector. Schematic diagram of the spectrophotometer is shown in Fig. 14.

Two light sources are focused on the diffraction grating that splits the light into its component wavelength as prism does. The purpose of the slit is to control the intensity of the incident light on the rotating disc that consists of three sections: a transparent section, a mirrored section and a black section. When light incidents on the disc, the following things may happen:

(A) If it comes to the transparent portion, it shall go straight and pass through the sample cell. It is then reflected by a mirror onto a 2nd rotating disc. This 2nd disc is rotating in such a way that when the beam of light comes from the 1st disc, it meets the mirrored part of the 2nd disc and gets reflected onto the detector as shown with the red path.

(B) If the light beam is incident on the mirrored section of the 1st rotating disc, it goes along the green path. After reflecting on the mirror, it passes through a reference cell. Then the light gets to the 2nd disc which rotates in such a way that the beam faces the transparent part. Hence it goes straight through to the detector and computer.

(C) If the light is incident on the first disc at the black section, it is fully blocked—and so for a short time no light passes through the UV-Vis spectrometer. This helps the computer to make allowance for current (if any) generated by the detector in absence of light. After rotating disc, the beam is allowed to pass through two cells, called as reference cell and sample cell. Reference cell

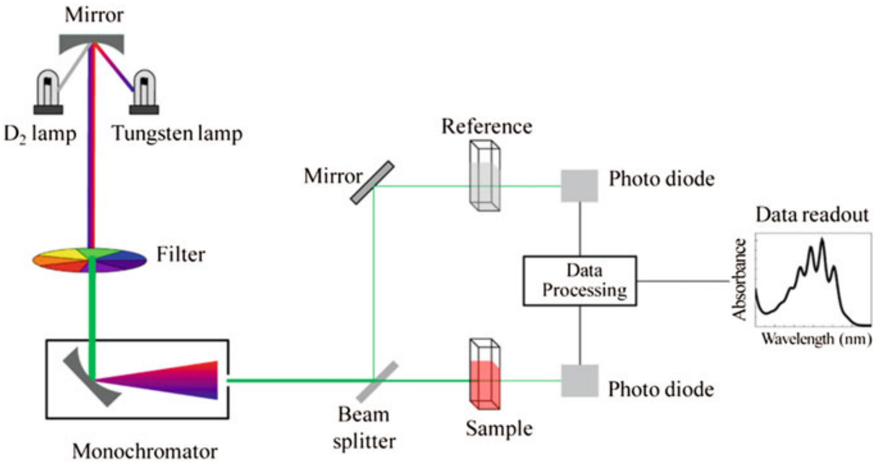


Fig. 14 Schematic of UV-visible spectrophotometer (*Source* www.wikipedia.org)

> contains either pure bare solvent (in case of liquid sample) or bare substrate (in case of film), whereas sample cell consists of sample. At the end of the measurement process, detector converts the incoming light beam into an electrical signal. The larger the electrical signal, the greater will be the light intensity.

Initially for each wavelength of light passing through the spectrometer, the intensity (I_0) of the light passing through the reference cell is recorded, called initialization process. Next step includes measurement of the intensity of the light beam passing through the cell for same wavelengths. This intensity is referred as I. Two different situations arise in this case. For larger wavelength having energy less than the band gap of the corresponding material, kept at reference cell, no light is absorbed by the sample ($I \approx I_0$). On the other hand, at lower wavelength with energy greater than the band gap, light will be absorbed by the sample, hence $I < I_0$. Therefore, using the following equation, absorbance (A) of the sample can easily be obtained,

$$A = \log_{10}\frac{I_0}{I}$$

Thus using this procedure, it is possible to determine the A in the whole range of the wavelength. If we have the knowledge of the depth of the material, then we may calculate the absorption coefficient (α) of the material. The relation between α and energy of the incident beam (h_v) was deduced by Tauc and is given by,

$$(\alpha h_v)^{1/n} = A\left(h_v - E_g\right)$$

here, E_g and A are the band gap of the corresponding materials and proportionality constant. 'n' assumes values like 2 and 1/2 in accordance with indirect and direct nature of electronic transition. According to the above equation, the plot of $(\alpha h_v)^{1/n}$ versus h_v must be a straight and from its intercept on the h_v axis, it is possible to determine the band gap of the material.

2.8 Optical Characterization by Ellipsometer

Ellipsometer is generally used to measure the thickness and to determine the optical properties like refractive indices, extinction coefficient, real and imaginary dielectric constant etc. of materials [25]. For films, colors are observed due to interference of light that is used to measure thickness of transparent bodies and this was first proposed by Brewster. In the presence of polarized light, Drude observed phase shift between them after reflection from film. The basic principle in this spectroscopic technique is that the change of polarization of plane polarized light after reflection from sample surfaces (mainly in thin film form), usually at the boundaries in planar multilayered materials. The system consists of light source, polarizer, monochromator, collimator, and detector. Light, from an arc lamp, is used to pass successively through a monochromator and a collimator those produce a narrow beam of light having specific wavelength. Then it passes through a polarizer (produces linearly polarized light) that functions on the basis of double refraction from uniaxial crystal. It is taken in form of special type of prism called Nicola Prism. Plane polarized light from polarizer incidents on the sample under study at various angle (varies between 50° and 80°) depending on the sample. After reflection from the sample, phase and amplitude of the both component of the plane polarized light changes i.e., reflected light is basically elliptically polarized that enters into the second polarizer, rotating in nature. Finally, light beam is collected at detector (Fig. 15).

The theory behind ellipsometer has been developed on the basis of Fresnel reflection at multilayer surfaces, starting from Maxwell's equation. The measurement is expressed in terms of ψ (psi) and Δ (delta), defined according to the relation, $\rho = \tan(\psi) \cdot \exp(i\Delta) = \frac{r_p}{r_s}$, where r_p and r_s are the complex Fresnel reflection coefficients for the samples for *p*—(in plane of incidence) and *s*—(perpendicular to the plane of incidence) polarized light. Actually, spectroscopic ellipsometer measures 'ρ' at various wavelengths. Since, ellipsometric measurement is related to phase of the light, so this spectroscopic is highly sensitive and requires no reference material. Even, the sensitivity is enhanced for multilayer thin film. A single wavelength ellipsometer produces two types of data ψ and Δ that allows two parameters of a material.

After acquiring the data, they are analyzed using the following procedures:

1. Modeling of the layers: the number of layers is chosen and basic structure concerning the contents of each layer is approximated. One must decide if any layer is anisotropic at this stage, and whether or not interface layers are to be modeled as a single effective medium approximation, or more complicated graded interface.
2. Assignment of the optical functions of each layer: Optical constants corresponding to each layer is considered. Here, the use existing sets of data or some other kind of parameterization is decided. For example, if the sample consists of

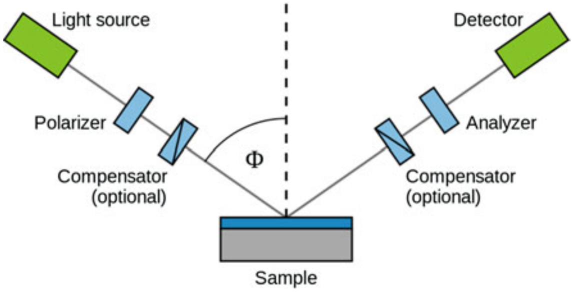


Fig. 15 Schematic setup of an ellipsometry experiment (*Source* Wikipedia under creative commons attribution ShareAlike3.0 license)

SiO_2 thin film grown on silicon substrate, it is very likely that the optical functions of the substrate layer will be very close to the reference values of crystalline silicon. The system may be modeled with three layers and a substrate: surface roughness/amorphous SiO_2/interface layer/crystalline silicon. The surface roughness (the top layer) may also be best modeled by Bruggeman effective medium consisting of 50 % air and 50 % SiO_2 (other % is also possible), while the middle layer is best modeled using a single Lorentz oscillator. The interface layer may be again be modeled using 2-medium Bruggeman effective medium that consists of 50 % layer 2 and 50 % silicon (% may be varied). The initial values for all parameters and film thicknesses must be assigned and the "floating" parameters selected.

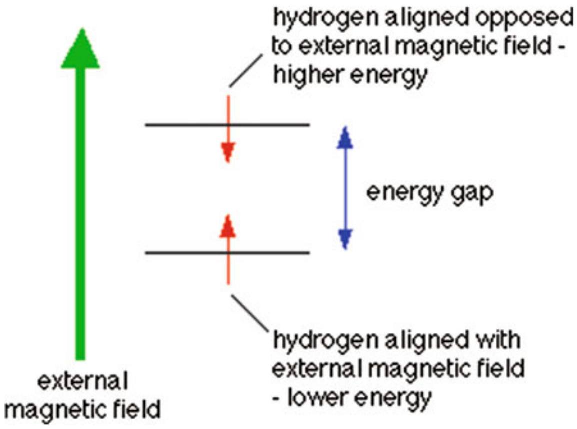
3. Fitting the data: The last step consists of the fitting process itself. Here, a metric is defined to distinguish good and bad fitting of data. Moreover, this metric, at least in a qualitative sense, also predicts the error limits of data. Regression technique is applied to tally whether there is any mismatch between the experimental data and data from model.

There are few models like Lorentz oscillator model, effective medium approximation model those are used in ellipsometric data fitting. Lorentz oscillator model being the oldest model is applicable for photon energy less than the band gap of the material. From this model dielectric constant (ε) and extinction coefficient (k) as function of photon energy (wavelength) is obtained. Effective medium approximation (EMA) model relies on the optical constants of thin films using an average of two or more other sets of optical functions. Three EMA models, namely Lorentz–Lorentz, Maxwell–Garnett and Buggeman models have been developed. Lorentz–Lorentz model is used for host materials with dielectric constant equal to one. Maxwell–Garnett model includes host materials with highest dielectric constant. Buggeman model is applicable when host medium possesses dielectric constant equal to the EMA dielectric constant. In this approximation, surface roughness is modeled by considering 50 % of void and 50 % of the underneath material.

2.9 Nuclear Magnetic Resonance

As it is named, nuclear magnetic resonance (NMR) deals with magnetic properties of nuclei. Generally, NMR deals with hydrogen nuclei, hence sometimes, it is also termed as proton NMR or 1H-NMR. [26] Hydrogen atoms those act as a little magnet can be aligned (having lower energy state) by the application of an external magnetic field. From general principle of physics, we know that there also exists another possibility. They may be aligned in the opposite direction of the field having higher energy states. Thus it is possible to reorient the alignment to a less stable state by supplying the sufficient amount of energy. The energy difference between two states depends on the strength of external magnetic field, but generally

Fig. 16 Schematic diagram for energy level splitting in the presence of magnetic field


hydrogen aligned opposed to external magnetic field - higher energy
energy gap
hydrogen aligned with external magnetic field - lower energy
external magnetic field

falls in the range of energies associated with radio waves, i.e., at around 60–100 MHz. But in the case of real materials, the effect of the surrounding electrons is to hinder the effect of the external magnetic field felt by the hydrogen nucleus. If the hydrogen atom is attached to an electronegative element, hindering effects become less effective and the external magnetic, needed to bring resonance, becomes smaller. Thus the resonating magnetic field represents the characteristic of its surrounding appears as a peak in the spectrum. The sizes of the two peaks give important information about the numbers of hydrogen atoms in each environment. The horizontal scale, presented in the unit of ppm, is termed as the chemical shift (δ). The scale has a zero point—at the right-hand end of the scale. The position of the peaks shows useful information about the environments to which the various hydrogen atoms belongs. The zero is where a peak is found due to the hydrogen atoms in tetramethylsilane—usually called TMS. Everything else is compared with this. NMR spectrum is generally taken after dissolving the material into some solvent. A solvent like carbon tetrachloride (CCl_4) can be used, which doesn't contain hydrogen or a solvent in which hydrogen atoms are replaced by its isotope e.g., deuterium may also be used. Deuterium atoms have sufficiently different magnetic properties from normal hydrogen as they don't produce any peak in the whole range of spectrum (Fig. 16).

References

1. C. Suryanarayana, Mechanical alloying and milling. Prog. Mater. Sci. **1–184**, 46 (2001)
2. S. Morup, J.Z. Jiang, F. Bodker, A. Horsewell, Crystal growth and the steady-state grain size during high energy ball-milling. Europhys. Lett. **441**, 56 (2001)
3. G. Scholz, R. Konig, J. Petersen, B. Angelow, I. Dorfel, E. Kemnitz, Chem. Mater. **20**, 5406 (2008)
4. G. Cao, *Nanostructures and Nanomaterials Synthesis, Properties* (Imperial College Press, London, 2004)

5. B.D. Yao, Y.F. Chan, N. Wang, Appl. Phys. Lett. **81**, 757 (2002). doi:10.1063/1.1495878
6. M.D. Shirk, P.A. Molian, A review of ultrashort pulsed laser ablation of materials. J. Laser Appl. **10**, 18 (1998)
7. M.O. Thompson, 328 Bard Hall Ithaca, NY 14853. (Personal communication)
8. A.Y. Cho, J.R. Arthur, Molecular beam epitaxy. Prog. Solid State Chem. **10**, 157–191 (1975)
9. A.N. Banerjee, C.K. Ghosh, K.K. Chattopadhyay, A.K. Hideki Minoura, A. Sarkar, A. Akiba, A. Kamiya, T. Endo, Low-temperature deposition of ZnO thin films on PET and glass substrates by DC-sputtering technique. Thin Solid Films **496**, 112–116 (2006)
10. S. Nandy, U.N. Maiti, C.K. Ghosh, K.K. Chattopadhyay, Enhancement p-type conductivity and band gap narrowing in heavily Al doped NiO thin films deposited by RF magnetron sputtering. J. Phys. Condens. Matt. **21**, 115804 (2009)
11. K. Minegishi, Y. Koiwai, Y. Kikuchi, K. Yano, M. Kasuga, A. Shimizu, Growth of p-type zinc oxide films by chemical vapor deposition. Jpn. J. Appl. Phys. **1453**, 36 (1997)
12. M. Leskela, M. Ritala, Atomic layer deposition chemistry: recent developments and future challenges. Angew. Chem. Int. Ed. **42**, 5548–5554 (2003)
13. C.J. Brinker, A.J. Hurd, G.C. Frye, P.R. Shunk, Ashley, sol-get thin-film formation. J. Ceram. Soc. Jpn. Int. Ed. **99**, 843 (1991)
14. J. Cheng, D.B. Fan, H. Wang, B.W. Liu, Y.C. Zhang, H. Yan, Chemical bath deposition of crystalline ZnS thin films. Semicond. Sci. Technol. **676**, 5 (2003)
15. C.B. Murray, C.R. Kagan, M.G. Bawendi, Annu. Rev. Mater. Sci. **30**, 545–610 (2000)
16. D. Sarkar, C.K. Ghosh, K.K. Chattopadhyay, Morphology control of rutile TiO_2 hierarchical architectures and their excellent field emission properties. Cryst. Eng. Comm. **24**, 2683 (2012)
17. F. Cerrina, Recent advances in X-ray lithography. J. Jpn. Appl. Phys. **4178**, 31 (1992)
18. E.B. Kley, Continuous profile writing by electron and optical lithography. Microelectron. Eng. **34**, 261–298 (1997)
19. F. Watt, A.A. Bettiol, J.A.V. Kan, E.J. Teo, M.B.H. Breese, Ion beam lithography and nanofabrication: a review. Int. J. Nanosci. **269**, 04 (2005)
20. B.D. Cullity, *Elements of X-Ray Diffraction* (Addison Wesley Publishing Company Inc., Boston)
21. L.M. Watt, *The Principles and Practice of Electron Microscopy* (Cambridge University Press, New York, 1985)
22. R.D. Heidenreich, *Fundamentals of Transmission Electron Microscopy* (Interscience Publisher, New York, 1964)
23. Y. Kuk, P.J. Silverman, Scanning tunneling microscope instruments. Rev. Sci. Instrum. **165**, 60 (1989)
24. J.L. Hutter, J. Bechhoefer, Calibration of atomic-force microscope tips. Rev. Sci. Instrum. **1868**, 64 (1993)
25. C.C. Kim, J.W. Garland, H. Abad, P.M. Raccah, Modeling the optical dielectric function of semiconductors: extension of the critical-point parabolic-band approximation. Phys. Rev. B **11749**, 45 (1992)
26. P.A. Bottomley, NMR imaging techniques and applications: a review. Rev. Sci. Instruments **1319**, 53 (1982)

Nanoscale MOSFET: MOS Transistor as Basic Building Block

Soumya Pandit

1 Introduction

Metal–Oxide–Semiconductor Field-Effect Transistor (MOSFET) has emerged over the last few decades as the basic building block of almost all computing devices. This steady growth of MOS transistors is attributed to the scaling of the underlying MOS technology which at present has reached the nanoscale (sub-90 nm) regime. Although the industry roadmap has suggested that the MOS transistor in its classical form is rapidly approaching toward some fundamental physical limits, yet many semiconductor industries prefer to use classical MOS transistor device structure because of the simplicity of operation and success of the use of such transistors in low cost-integrated circuits and systems. This chapter attempts to provide an introductory overview of the structure and principle of operation of classical MOS transistor. The equivalent circuit models of MOS transistor for analog and digital applications are also discussed. Finally, a basic idea of compact device modeling and the use of such models in circuit simulation process are introduced.

2 Basic Structure of MOS Transistor

The MOSFET is a four terminal device. These four terminals are source, drain, gate, and body/bulk/substrate. Figure 1a shows the cross-sectional view of the structure of an n-channel enhancement-type MOSFET. The transistor is fabricated on a p-type semiconductor substrate. Two heavily doped n-type regions are developed in the substrate. One of these regions with metal contacts above it forms the source

---

S. Pandit (✉)
Institute of Radio Physics and Electronics, Calcutta University, Kolkata, India
e-mail: soumya.pandit.rpe@gmail.com

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_7

(S) terminal and another one forms the drain (D) terminal. A thin layer of silicon dioxide (SiO_2) is grown above the semiconductor surface, covering the area between the source and the drain regions. On top of the oxide layer, a metal layer is deposited to form the gate (G) electrode of the device. The semiconductor region underneath the oxide (between the source and the drain terminals) is referred to as the channel region. The fourth terminal of the device is body (B), which is formed by constructing a metal contact at the substrate. For a p-channel enhancement-type MOSFET, the transistor is developed on an n-type substrate. The cross-sectional view of a p-channel MOS transistor is illustrated in Fig. 1b. It may be noted that the present-day very large-scale integration (VLSI) circuits use enhancement-type transistors. However, there exists another type of MOS transistors, known as depletion-type MOS transistor. The physical construction of a depletion-type MOS transistor is identical to an enhancement-type MOS transistor, with the difference that in the depletion-type MOS transistor, the conducting channel between the source and the drain region is pre-implanted. However, for enhancement-type MOS transistor, it is induced by an externally applied gate voltage. In the rest of the chapter, we will concentrate only on enhancement-type MOS transistor.

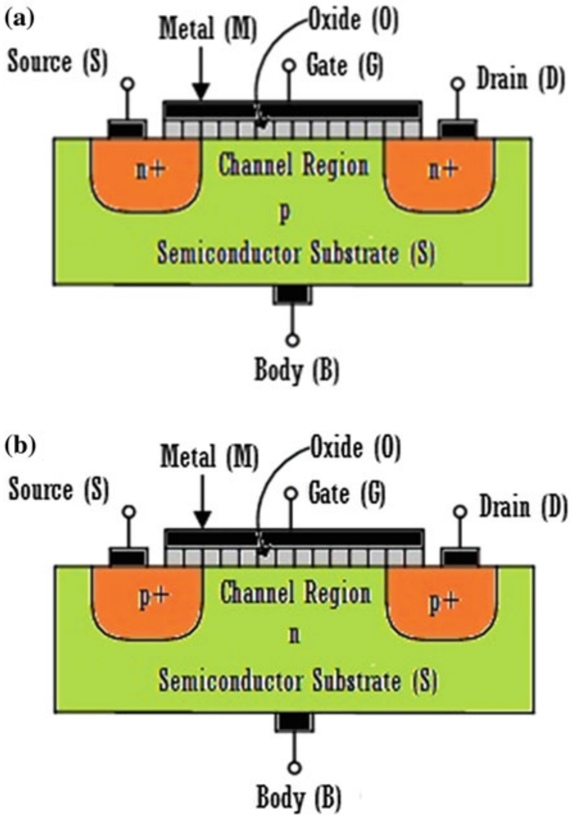


Fig. 1 Cross-sectional view of the structure of **a** n-channel MOS transistor and **b** p-channel MOS transistor

3 Electrical Characteristics of MOS Transistor

The electrical characteristics of MOS transistor are described by the flow of current carriers between the source and the drain terminals with the application of voltage at the gate and the drain terminals with respect to a reference terminal, which usually is the source terminal. This section discusses the generation and transport of charges through the channel region as function of these terminal voltages.

3.1 Generation of Charges in the Channel Region

The generation of charges in the channel region of an MOS transistor is best understood through the concept of MOS capacitor. The MOS capacitor consists of a metal–oxide–semiconductor structure as illustrated in Fig. 2. An external voltage

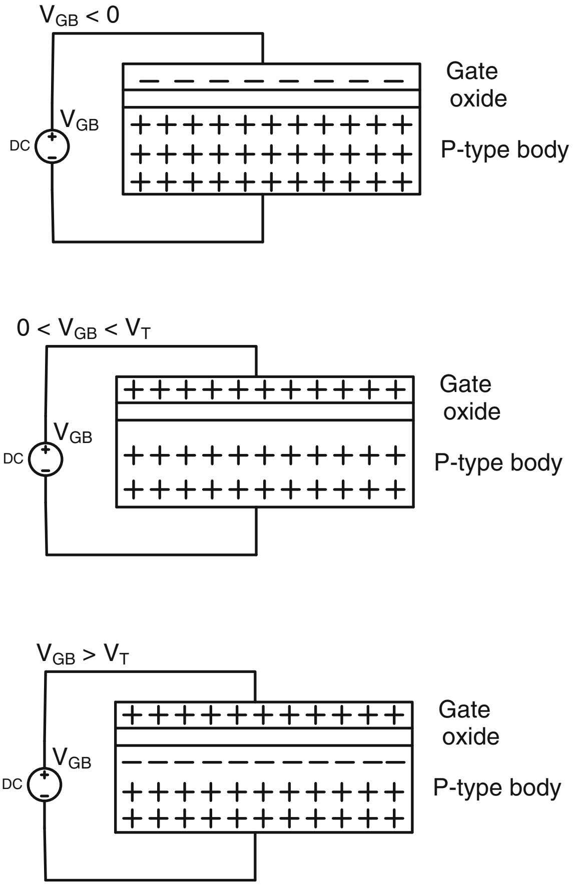


Fig. 2 MOS capacitor structure demonstrating **a** accumulation, **b** depletion, and **c** inversion region

V_{GB} is applied between the gate and the body of the transistor. Three cases are identified, which are discussed below.

Case 1. Let us first consider the case when $V_{\mathrm{GB}} < 0$ (Fig. 2a). This effectively deposits negative charges on the metal. Consequently, equal amount of positive charges are accumulated at the surface of the semiconductor. For p-type substrate, this occurs by attraction of mobile holes from the substrate.

Case 2. The second case is when a small positive voltage $V_{\mathrm{GB}} > 0$ is applied at the gate (Fig. 2b). The positive voltage deposits positive charges on the metal and demands negative charges at the surface of the semiconductor. For p-type semiconductor, these arise from depletion of holes from the region near the surface, leaving behind uncompensated ionized acceptors.

Case 3. The third case is when the applied voltage V_{GB} is sufficiently positive ($V_{\mathrm{GB}} > V_{\mathrm{T}}$), so that a large number of electrons from the deep substrate are attracted under the gate oxide (Fig. 2c). For an MOS capacitor, these electrons come from the process of electron–hole generation in the depletion region, caused by thermal vibration of the lattice. The semiconductor surface under this condition, is said to be inverted, i.e., no longer p-type. This is referred to as the case of inversion. The conductive channel of electrons in the p-type body, underneath the gate oxide between the source and drain regions is referred to as the inversion layer. For a p-type substrate, when the semiconductor surface becomes as strongly n-type as the substrate is p-type, strong inversion is said to occur in the channel region. It may be noted that this n-type surface layer is formed not by doping, but by inversion of the originally p-type semiconductor with the application of external gate voltage. This inverted layer is the key to MOS transistor operation.

If the substrate is made of n-type material, the above picture should be modified in a rather obvious manner. The inversion layer will consist of holes which will be attracted to the surface if V_{GB} is sufficiently negative. On the other hand, if V_{GB} is sufficiently positive, electrons will pile up at the surface.

The minimum gate voltage required to produce an inversion layer in the channel is called the threshold voltage V_{T}. An expression for the threshold voltage V_{T} is written as

$$V_{\mathrm{T}} = V_{\mathrm{FB}} + 2\phi_{\mathrm{F}} - \frac{Q_{\mathrm{b}}}{C_{\mathrm{ox}}} \tag{1}$$

where C_{ox} is the oxide capacitance per unit area, Q_{b} is the charge per unit area due to uncovered atoms in the depletion region, and V_{FB} is referred to as the flat-band voltage. The three terms in the right-hand side of (1) are discussed below separately.

The flat-band voltage is given as

$$V_{FB} = \Phi_{MS} - \frac{Q_i}{C_{ox}} \tag{2}$$

where Φ_{MS} is referred to as the metal–semiconductor work function. This is interpreted as the external voltage needs to be applied externally at the gate terminal to precisely cancel the contact potential difference between the semiconductor substrate material and gate material. In (2), Q_i represents the effective interface charge per unit area present at the oxide–semiconductor interface. This charge is almost always positive for both p- and n-type substrates. Therefore, the flat-band voltage is the external voltage that needs to be applied between the gate and the substrate terminals to keep the semiconductor everywhere neutral by canceling the effects of the contact potential and the effective interface charge.

The surface potential ψ_s is the total potential drop, defined from the semiconductor surface to a point in the bulk region. At the onset of strong inversion in the channel region, the surface potential is calculated to be equal to $2\phi_F$, i.e., $\psi_s = 2\phi_F$, where ϕ_F is the difference between the Fermi potential and the intrinsic potential in the bulk semiconductor. In between, ϕ_F and $2\phi_F$, the channel remains in the weak inversion region. It is to be noted that the surface potential value is positive for n-channel MOS transistor and negative for p-channel MOS transistor.

An external voltage is required at the gate terminal to sustain the depletion layer charge density Q_b. This is given as $\frac{-Q_b}{C_{ox}}$. The quantity Q_b is negative for n-channel MOS transistor and is positive for p-channel MOS transistor.

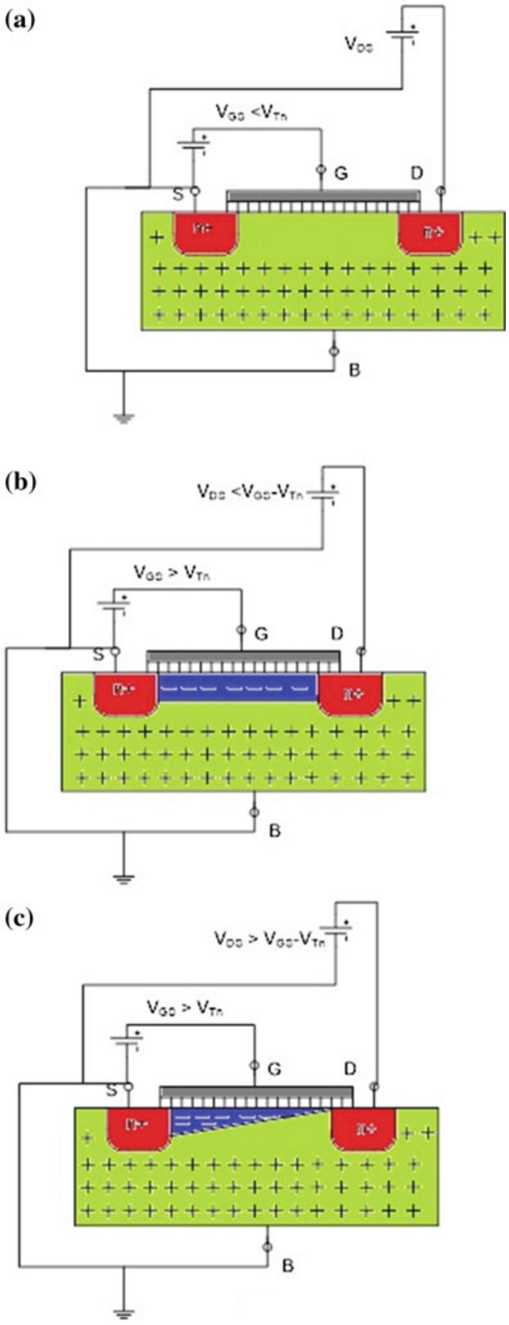
3.2 Transportation of Charges in the Channel Region—MOS Transistor Action

In order to understand the mechanism of transistor action, let us consider an n-channel enhancement-mode transistor. The substrate is made of p-type semiconductor. The source and the body terminals are connected to the ground, i.e., $V_S = 0$ and $V_B = 0$.

In Fig. 3a, the gate-source voltage V_{GS} is less than the threshold voltage V_{Tn}. Between the drain and the source terminals, a potential V_{DS} is applied. The body, which is heavily doped p-type semiconductor, has sufficient number of holes, but no free electrons. On the other hand, the source and the drain regions have enough number of free electrons. With the absence of conductive channel between the source and the drain regions, negligible amount of current flows. This mode of transistor operation is called *cutoff*.

In Fig. 3b, the gate voltage is greater than the threshold voltage V_{Tn}. Now an inversion region, consisting of free electrons, is created in the semiconductor surface underneath the gate oxide, connecting the source and the drain regions. With this, the structure looks electrically like an induced n-type resistor. The number of

Fig. 3 MOS transistor operation: **a** cutoff, **b** linear, and **c** saturation



carriers in the channel region increases with the gate voltage, thereby increasing the conductivity of the channel region. When the applied drain–source potential V_{DS} is greater than zero, the horizontal electric field helps drift motion of the free electrons from the source to the drain region. Thus current I_{DS} starts flowing from the drain to the source terminal. This mode of operation is termed as *linear*; the drain current increases linearly with the applied drain potential.

With the flow of drain current, an ohmic voltage drop occurs along the channel. Therefore, the voltage difference between the gate and the channel reduces from V_G near the source to $V_{GD} = (V_G - V_D)$ near the drain end. If the drain voltage V_{DS} is sufficiently high such that $V_{GD} < V_{Tn}$, the inversion charge density at the drain terminal becomes zero, so that the surface channel vanishes at the drain end of the channel. The channel is said to be "*pinched off*," which is illustrated in Fig. 3c. Under such conditions, electrons in the channel are pulled into the pinch-off region and travel at the saturation drift velocity due to the very high longitudinal electric field along the channel. This mode of operation is referred to as *saturation*, because the drain current does not increase significantly with drain potential.

3.3 Derivation of I–V Model

Let us consider an n-channel MOS transistor as shown in Fig. 4. The source is assumed grounded and external bias voltages V_{GS} and V_{DS} are applied. If the applied gate-source voltage V_{GS} is greater than the threshold voltage V_{Tn}, a conducting channel is created. The conductivity of the channel is controlled by the vertical electric field due to $(V_{GS} - V_{Tn})$. The current that flows from the drain to source is determined by the horizontal electric field due to V_{DS}. The exact value of the drain current I_{DS} is determined by both the vertical and horizontal electric field, thus explaining the term field-effect transistor.

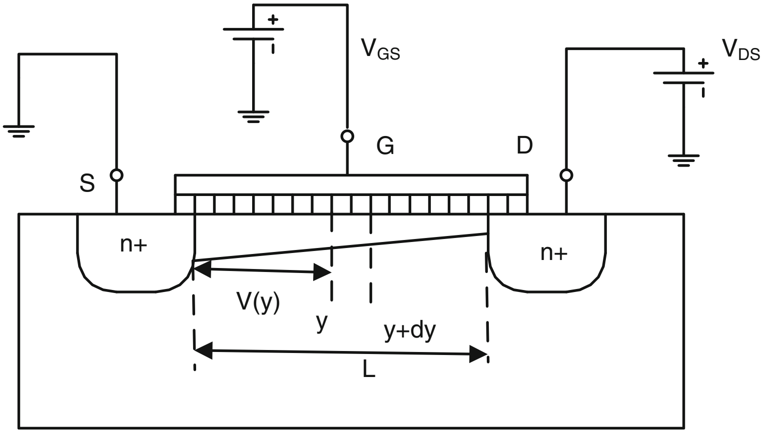


Fig. 4 NMOS transistor with bias voltages applied

The drain-to-source current is

$$I_{\mathrm{DS}} = WQ_{\mathrm{n}}(y)v_{\mathrm{d}}(y) \tag{3}$$

where $Q_{\mathrm{n}}(y)$ is the inversion charge per unit area at a position y in the channel and $v_{\mathrm{d}}(y)$ is the drift velocity of carriers at that position. Here, $y = 0$ represents the source end and $y = L$ represents the drain end. In (3), W represents the width of the channel region.

$$v_{\mathrm{d}}(y) = \mu_{\mathrm{ns}}\xi(y) \tag{4}$$

where μ_{ns} is the mobility of electrons at the channel surface and $\xi(y)$ is the electric field. It is given as

$$\xi(y) = -\frac{\mathrm{d}V_{\mathrm{CS}}(y)}{\mathrm{d}y} \tag{5}$$

$V_{\mathrm{CS}}(y)$ is the potential in the channel at a point y along the channel length, measured w.r.t the source. Therefore, we get

$$I_{\mathrm{DS}} = -WQ_{\mathrm{n}}(y)\mu_{\mathrm{ns}}\frac{\mathrm{d}V_{\mathrm{CS}}(y)}{\mathrm{d}y} \tag{6}$$

At this point, we make three assumptions in order to make the I-V model simple. First, the mobility μ_{ns} of the carriers in the channel region is constant. Second is the gradual channel approximation, which assumes that the variation of the electric field along the channel direction is much less than the corresponding variation along the direction perpendicular to the channel. Third, the threshold voltage is not a function of the position y along the channel.

With these assumptions, the inversion charge density is given by

$$Q_{\mathrm{n}}(y) = -C_{\mathrm{ox}}(V_{\mathrm{GS}} - V_{\mathrm{Tn}} - V_{\mathrm{CS}}(y)) \tag{7}$$

Substituting (7) in (6) and integrating within appropriate limits, we get the following

$$\int_0^L I_{\mathrm{DS}}\mathrm{d}y = \int_0^{V_{\mathrm{DS}}} W\mu_{\mathrm{ns}}C_{\mathrm{ox}}[V_{\mathrm{GS}} - V_{\mathrm{Tn}} - V_{\mathrm{CS}}(y)]\mathrm{d}V_{\mathrm{CS}} \tag{8}$$

Carrying out the integration, we get

$$I_{\mathrm{DS}} = k_{\mathrm{n}}\frac{W}{L}\left[(V_{\mathrm{GS}} - V_{\mathrm{Tn}})V_{\mathrm{DS}} - \frac{1}{2}V_{\mathrm{DS}}^2\right] \tag{9}$$

where $k_{\mathrm{n}} = \mu_{\mathrm{ns}}C_{\mathrm{ox}}$. In circuit design, it is more convenient to express the I-V equations in terms of electrical rather than physical parameters. For this reason, the drain current is expressed as

$$I_{\mathrm{DS}} = \beta_{\mathrm{n}}\left[(V_{\mathrm{GS}} - V_{\mathrm{Tn}})V_{\mathrm{DS}} - \frac{1}{2}V_{\mathrm{DS}}^2\right] \tag{10}$$

where the process transconductance parameter β_{n} is given in terms of physical parameters as

$$\beta_{\mathrm{n}} = \mu_{\mathrm{ns}} C_{\mathrm{ox}} \frac{W}{L} \tag{11}$$

3.4 Modes of Operation of MOS Transistor

The various modes of operation of an MOS transistor have already been discussed qualitatively. Here we present the quantitative description in terms of *I*-*V* equation for various modes. These depend upon the magnitude of the overdrive voltage $(V_{\mathrm{GS}} - V_{\mathrm{Tn}})$.

1. Cutoff Mode: If $(V_{\mathrm{GS}} - V_{\mathrm{Tn}}) \le 0$, then the MOS transistor operates in the cutoff region. In this region,

   $$I_{\mathrm{DS}} = 0 \quad (V_{\mathrm{GS}} - V_{\mathrm{Tn}}) \le 0 \tag{12}$$

   In this mode, the channel acts like an open circuit
2. Linear Mode: This region is defined as

   $$I_{\mathrm{DS}} = \beta_{\mathrm{n}}\left[(V_{\mathrm{GS}} - V_{\mathrm{Tn}})V_{\mathrm{DS}} - \frac{1}{2}V_{\mathrm{DS}}^2\right] \tag{13}$$

   The condition is

   $$V_{\mathrm{Tn}} < V_{\mathrm{GS}}, \quad 0 < V_{\mathrm{DS}} \le (V_{\mathrm{GS}} - V_{\mathrm{Tn}}) \tag{14}$$

   If V_{DS} is small enough so that the term $\frac{1}{2}V_{\mathrm{DS}}^2$ can be neglected in the drain current expression, we get

   $$I_{\mathrm{DS}} = \beta_{\mathrm{n}}(V_{\mathrm{GS}} - V_{\mathrm{Tn}})V_{\mathrm{DS}} \tag{15}$$

   It is thus observed that the drain current is linearly proportional to the output drain–source voltage. This implies that the path from the source to the drain can be represented by a linear resistor equal to

   $$R_{\mathrm{on}} = \frac{1}{\beta_{\mathrm{n}}(V_{\mathrm{GS}} - V_{\mathrm{Tn}})} \tag{16}$$

We thus conclude that an MOS transistor acts as a linear resistor, controlled by the overdrive voltage ($V_{GS} - V_{Tn}$) provided $V_{DS} \ll 2\ (V_{GS} - V_{Tn})$

3. Saturation Mode: This mode is defined as

$$\begin{aligned}\frac{dI_{DS}}{dV_{DS}} &= 0 \Rightarrow V_{DS}(\text{sat}) \\ &= (V_{GS} - V_{Tn})\end{aligned} \tag{17}$$

The drain–source voltage $V_{DS}(\text{sat})$ is referred to as the saturation voltage. The drain current in this mode of operation is written by substituting (17) in (10) which is as follows:

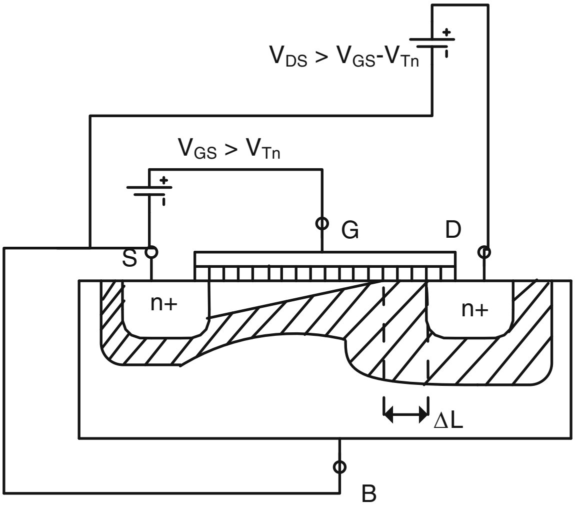
$$I_{DS} = \frac{\beta_n}{2}(V_{GS} - V_{Tn})^2 \qquad 0 < (V_{GS} - V_{Tn}) \leq V_{DS} \tag{18}$$

It is to be noted that the above *I*–*V* equations are for n-channel MOS transistor. The *I*–*V* equations of a p-channel MOS transistor are essentially same as that of the n-channel transistor, except that all voltage and current polarities are reversed.

3.5 Channel Length Modulation

It is observed from (18) that the drain current remains constant once V_{DS} is greater than the saturation voltage $V_{DS}(\text{sat})$. However, in reality this is not true. As the drain–source voltage is increased beyond the saturation voltage, the drain current increases slightly. This phenomenon is referred to as the channel length modulation effect. Let us consider Fig. 5. As the drain voltage increases beyond the saturation

Fig. 5 Channel length modulation phenomenon illustrated


V_{DS} > V_{GS}-V_{Tn}
+
V_{GS} > V_{Tn}
+
G
D
S
n+
n+
ΔL
B

voltage V_{DS}(sat), the pinch-off point where the surface channel collapses begins to move slightly toward the source as shown in Fig. 5. The distance between the pinch-off point and the drain, ΔL, is referred to as the amount of channel length modulation by the drain voltage. The drain current is then obtained by simply replacing L with $L - \Delta L$ in (18). Thus the channel length is effectively reduced and hence I_{DS} increases. The channel length modulation effect is incorporated in the *I*–*V* equation by multiplying the drain current expression under saturation condition with the factor ($1 + \lambda V_{DS}$), where λ is called the channel length modulation coefficient.

$$I_{DS} = \frac{\beta_n}{2}(V_{GS} - V_{Tn})^2(1 + \lambda V_{DS}) \quad 0 < (V_{GS} - V_{Tn}) \le V_{DS} \tag{19}$$

The coefficient λ is related to the channel length reduction as $\lambda V_{DS} = \frac{\Delta L}{L}$. It is to be noted that the reciprocal of λ is sometimes referred to as the Early voltage V_A. Typical values of λ are in the range of 0.05–0.005 V^{-1}.

3.6 Body Effect

All MOS transistors are generally fabricated on a common substrate and the substrate (body) voltage of all the transistors is normally constant. When a circuit is made up of several numbers of transistors, it often happens that several transistors are connected in series. This results in different source-to-substrate potentials for the different transistors. This is illustrated in Fig. 6. For the transistors M2 and M3, the source and substrate potentials are different.

In order to explain the effect of substrate bias on the channel inversion charge, consider application of a voltage $V_{SB} > 0$ ($V_{BS} < 0$) such that the n^+ (source)-p (body)

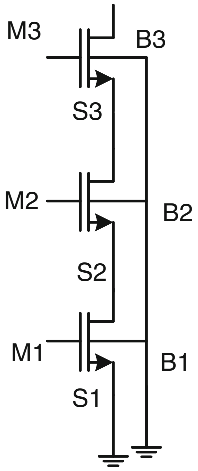


Fig. 6 Introduction of body effect in transistor M2 and M3

junction is reverse biased. As a result, the depletion region under the strongly inverted surface is widened. More acceptor atoms will be uncovered and the total depletion region charge will therefore increase. If the gate voltage is kept constant, with the availability of more ionized acceptor atoms, fewer free electrons are needed in the inversion layer. The level of inversion thereby reduces. Higher gate voltage is needed to restore the original level of inversion. Therefore, the threshold voltage increases. This increase of threshold voltage due to nonzero source-to-substrate voltage is referred to as the body effect/substrate bias effect.

The depletion charge densities in presence and absence of substrate bias are given by

$$Q_\mathrm{b} = \sqrt{2qN_\mathrm{A}\varepsilon_\mathrm{Si}(2\Phi_\mathrm{F} - V_\mathrm{BS})} \tag{20}$$

$$Q_\mathrm{b0} = \sqrt{2qN_\mathrm{A}\varepsilon_\mathrm{Si}(2\Phi_\mathrm{F})} \tag{21}$$

where N_A is the acceptor doping concentration in the semiconductor substrate and ε_Si is the permittivity of Si. With these, the threshold voltage is given as

$$V_\mathrm{T} = V_\mathrm{T0} + \gamma\left(\sqrt{2\Phi_\mathrm{F} - V_\mathrm{BS}} - \sqrt{2\Phi_\mathrm{F}}\right) \tag{22}$$

Here V_T0 is the threshold voltage in absence of the substrate bias and is given as

$$V_\mathrm{T0} = V_\mathrm{FB} + 2\Phi_\mathrm{F} - \frac{Q_\mathrm{b0}}{C_\mathrm{ox}} \tag{23}$$

The body effect coefficient γ is defined as

$$\gamma = \frac{\sqrt{2q\varepsilon_\mathrm{Si}N_\mathrm{A}}}{C_\mathrm{ox}} \tag{24}$$

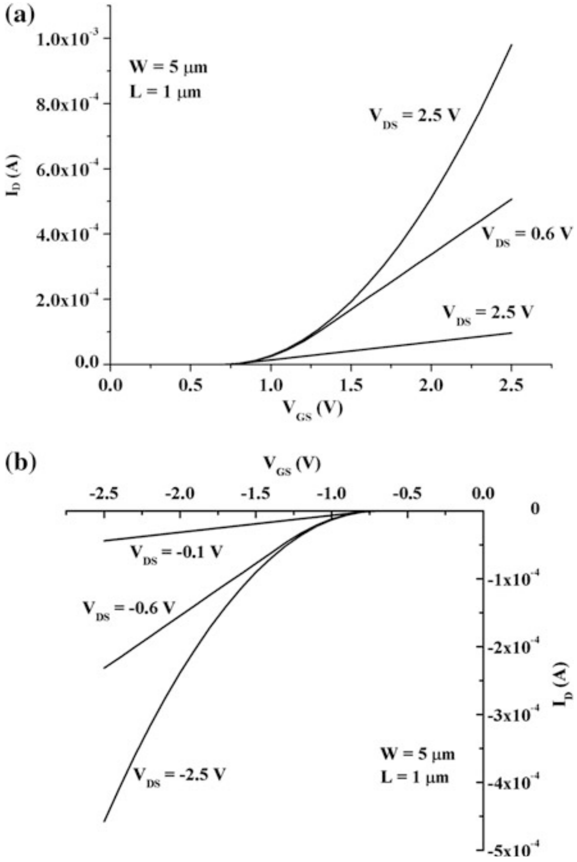
It follows from (24) that the body effect is stronger for heavier substrate doping and/or thicker oxides.

3.7 I-V Characteristic Curves

3.7.1 Input Characteristics

The input characteristics, depict the variations of the output drain current I_DS as function of the input gate-source voltage V_GS keeping the drain–source voltage V_DS constant. From this curve, threshold voltage of the MOS transistor can be estimated. Figure 7a and b show the input characteristics for an n-channel and a p-channel MOS transistor.

Fig. 7 Input characteristics: **a** n-channel MOS transistor and **b** p-channel MOS transistor



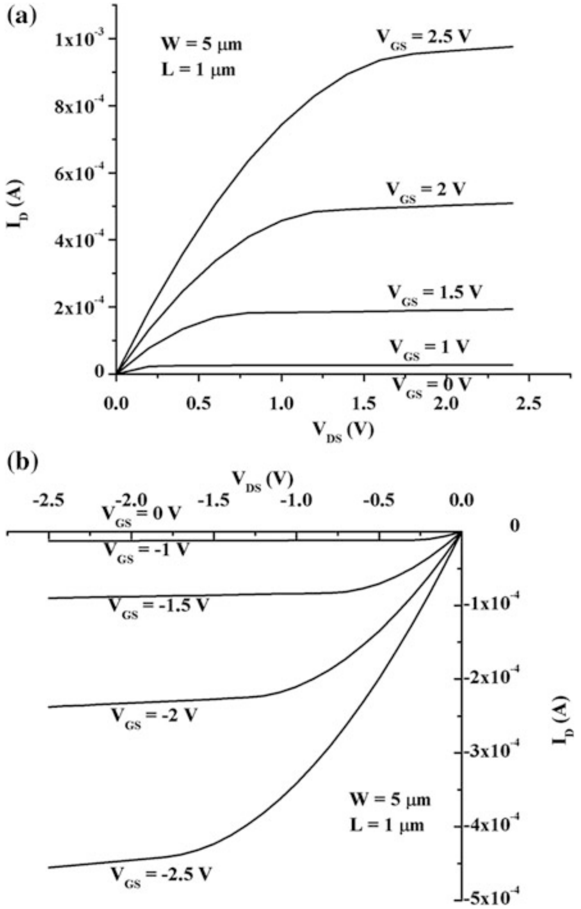
3.7.2 Output Characteristics

The output characteristics (see Fig. 8a, b) depict the variations of the output drain current I_{DS} as function of the drain–source voltage V_{DS} keeping the input gate-source voltage V_{GS} constant. From this curve, the three regions of the *I*–*V* characteristics as discussed earlier are observed.

It is observed that in the cutoff region, the drain current is almost zero. In the linear region, the drain current increases linearly with the drain–source voltage. On the other hand, in the saturation region, the drain current is nearly independent of the drain–source voltage. In addition, it is observed that in the saturation region, the drain current increases with the drain–source voltage to a certain extent. This is due to the channel length modulation phenomenon.

From both the input and output characteristics, it is evident that the directions of drain current flow for n-channel MOS and p-channel MOS transistor are opposite.

Fig. 8 Output characteristics: **a** n-channel MOS transistor and **b** p-channel transistor



4 Analog Equivalent Circuit

4.1 Low-Frequency Equivalent Circuit

In this subsection, we derive the low-frequency analog equivalent circuit model of the MOS transistor. By assuming low frequency, we neglect the effect of various MOS capacitors, assuming them to be open circuit. We also assume that the input signal amplitude is small such that the nonlinear characteristics of the device around the quiescent point can be approximated by linear characteristics. This considerably facilitates the design process, especially when several transistors are taken into account in the design process. As a thumb rule, the magnitude of the small signal amplitude is assumed to be much less than one-tenth of the magnitude of the corresponding large signal amplitude.

4.1.1 Derivation of the Model

From the previous discussion, the drain current of an MOS transistor is expressed through the following functional relationship

$$i_{\mathrm{D}} = f(v_{\mathrm{GS}}, v_{\mathrm{DS}}, v_{\mathrm{BS}}) \tag{25}$$

Making Taylor's series expansion of (25) around the quiescent point $Q(I_{\mathrm{DS}}$, V_{GS}, V_{DS}, $V_{\mathrm{BS}})$ and neglecting the higher order terms, we get

$$\Delta i_{\mathrm{D}} = \left.\frac{\partial i_{\mathrm{D}}}{\partial v_{\mathrm{GS}}}\right|_{Q} \Delta v_{\mathrm{GS}} + \left.\frac{\partial i_{\mathrm{D}}}{\partial v_{\mathrm{DS}}}\right|_{Q} \Delta v_{\mathrm{DS}} + \left.\frac{\partial i_{\mathrm{D}}}{\partial v_{\mathrm{BS}}}\right|_{Q} \Delta v_{\mathrm{BS}} \tag{26}$$

Using small signal notation, (26) may be written as follows:

$$i_{\mathrm{d}} = g_{\mathrm{m}} v_{\mathrm{gs}} + g_{\mathrm{ds}} v_{\mathrm{ds}} + g_{\mathrm{mb}} v_{\mathrm{bs}} \tag{27}$$

In (27), g_{m}, g_{ds}, and g_{mb} are the small signal parameters evaluated at the dc quiescent point Q. The values of these small signal parameters depend on the operating region. This is discussed as follows.

The small signal transconductance parameter in the saturation region is expressed as

$$g_{\mathrm{m}} = \sqrt{2I_{\mathrm{DS}}\beta_{\mathrm{n}}(1+\lambda V_{\mathrm{DS}})} \simeq \beta_{\mathrm{n}}(V_{\mathrm{GS}} - V_{\mathrm{T}}) = \sqrt{2\beta_{\mathrm{n}} I_{\mathrm{DS}}} \tag{28}$$

We observe that the transconductance of the MOS transistor is proportional to the square root of the dc bias current and depends on device geometry.

The small signal channel transconductance due to v_{BS} is defined as

$$g_{\mathrm{mb}} = \frac{\partial i_{\mathrm{D}}}{\partial v_{\mathrm{BS}}} = \frac{\partial i_{\mathrm{D}}}{\partial V_{\mathrm{T}}} \cdot \frac{\partial V_{\mathrm{T}}}{\partial v_{\mathrm{BS}}} = g_{\mathrm{m}} \cdot \frac{\gamma}{2\sqrt{2\Phi_{\mathrm{F}} - V_{\mathrm{BS}}}} = \eta g_{\mathrm{m}} \tag{29}$$

The factor η is typically in the range 0.1–0.3; therefore, the gate transconductance is typically a factor of 3–10 times larger than the body transconductance. The small signal channel conductance $g_{\mathrm{ds}}(g_0)$ is defined as

$$g_{\mathrm{ds}} = g_0 = \frac{I_{\mathrm{DS}}\lambda}{1+\lambda V_{\mathrm{DS}}} \simeq I_{\mathrm{DS}}\lambda = \frac{I_{\mathrm{DS}}}{V_{\mathrm{A}}} \tag{30}$$

It is observed from (30) that the channel conductance is dependent upon the channel length L through λ, which is inversely proportional to L.

The small signal parameters at the linear region are defined as

$$g_{\mathrm{m}} = \beta V_{\mathrm{DS}} \tag{31}$$

$$g_{\mathrm{mb}} = \eta g_{\mathrm{m}} \tag{32}$$

$$g_{\mathrm{ds}} = \beta(V_{\mathrm{GS}} - V_{\mathrm{T}} - V_{\mathrm{DS}}) \tag{33}$$

The dependence of the small signal parameters on the large signal quantities is emphasized.

4.1.2 Equivalent Circuit

The equivalent circuit representation corresponding to the model derived above is shown in Fig. 9. The equivalent circuit is adequate for most low-frequency small signal analyses. However, in practice, each terminal of the MOS transistor exhibits a finite ohmic resistance which occurs due to the resistivity of the material and the contacts. Proper layout strategies can minimize such resistances. These are therefore, not included in the equivalent circuit representation.

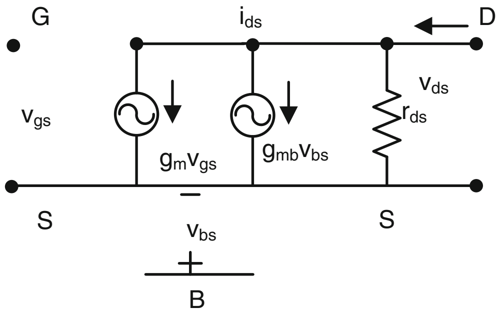
4.2 High-Frequency Equivalent Circuit

For high-frequency applications, the various intrinsic and extrinsic capacitances associated with MOS transistors need to be considered while developing the model. These are shown in Fig. 10. The capacitors of Fig. 10 can be separated into two broad types: (i) depletion capacitors and (ii) charge storage or parallel plate capacitors. These are discussed individually in details below.

4.2.1 Depletion Capacitors

The depletion capacitors are formed between the junctions of two oppositely doped semiconductors. For an n-channel MOS transistor, the n^+ doped source/drain region and p-doped substrate region forms a p-n junction. These p-n junctions should be

Fig. 9 Small signal equivalent circuit for low-frequency applications



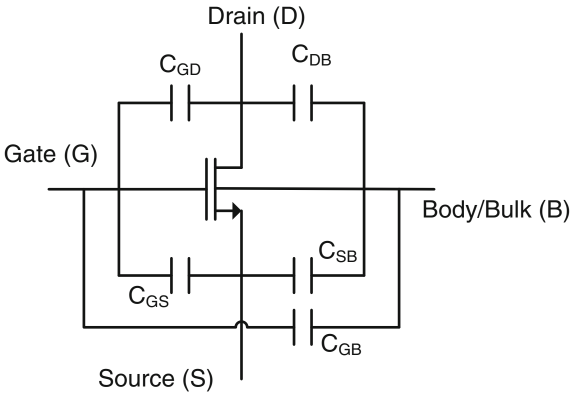


Fig. 10 Intrinsic and extrinsic capacitances associated with MOS transistor

reverse biased during normal operation and each junction exhibits a voltage-dependent parasitic capacitance associated with its depletion region. Thus depletion capacitors are formed between these two junctions. The capacitance values of the depletion capacitors are functions of the voltage across the junction. If the doping levels in the source, drain and body regions are assumed to be constant, then the depletion capacitances are expressed as

$$C_{\mathrm{XB}} = \frac{C_{\mathrm{j}} \cdot A_{\mathrm{X}}}{\left(1 + \frac{V_{\mathrm{XB}}}{\psi_0}\right)^{m_{\mathrm{j}}}} \tag{34}$$

where X = D for C_{DB} and X = S for C_{SB}. A_{X} is the area of the source or drain, C_{j} is the zero-bias junction capacitance per unit area, i.e., the capacitance when the voltage across the diode is zero. V_{XB} is the reverse bias voltage across the junction, ψ_0 is the built in junction potential, and m_{j} is the source/drain–substrate junction grading coefficient (1/2 or step junction and 1/3 for linearly graded junction). Physically, a depletion capacitor is like a tub, as seen from Fig. 11. It has a bottom with an area equal to the area of the drain or source. In addition, it has a sidewall which is also part of the depletion region. Thus A_{X} in (34) consists of both the bottom and sidewall assuming the zero-bias capacitances of the two regions are similar. Thus more accurately, the depletion capacitance is given as

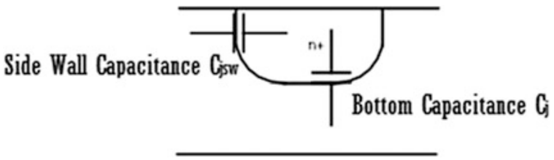


Fig. 11 Decomposition of depletion capacitors into bottom and sidewall component

$$C_{\mathrm{XB}} = \frac{C_{\mathrm{jb}} \cdot A_{\mathrm{X}}}{\left(1 + \frac{V_{\mathrm{XB}}}{\psi_0}\right)^{m_{\mathrm{j}}}} + \frac{C_{\mathrm{jsw}} \cdot P_{\mathrm{X}}}{\left(1 + \frac{V_{\mathrm{XB}}}{\psi_0}\right)^{m_{\mathrm{jsw}}}} \tag{35}$$

where P_{X} is the perimeter of the source or drain and m_{jsw} is the bulk-source/drain sidewall grading coefficient.

4.2.2 Charge Storage Capacitors

The charge storage capacitors of a MOS device are the gate-to-source capacitor C_{GS}, gate-to-drain capacitor C_{GD}, and the gate-to-bulk C_{GB} capacitor. Figure 12 shows a cross-section of the various capacitances that constitute the charge storage capacitors of MOS transistor. This consists of intrinsic capacitances and parasitic capacitances. The values of the intrinsic capacitors are bias dependent, while the values of the parasitic capacitors are bias independent.

The parasitic gate-to-source and the gate-to-drain capacitors, i.e., C_3 and C_4 are formed due to the overlap of the gate poly with the source and drain separated by a dielectric.

The cause of this overlapping is due to lateral diffusion of the source and drain regions.

Let the amount of overlap be designated by LD. The overlap capacitances are then approximated as

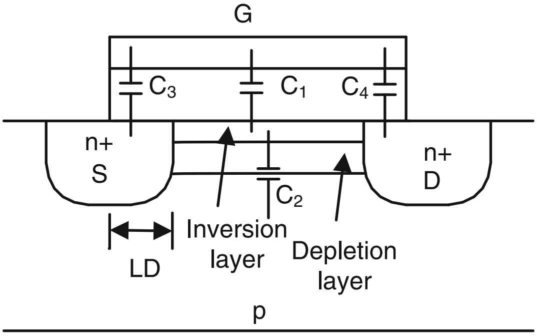
$$C_3 = C_4 \cong \mathrm{LD} \cdot W_{\mathrm{eff}} \cdot C_{\mathrm{ox}} = CGXO \cdot W_{\mathrm{eff}} \tag{36}$$

Here W_{eff} is the effective channel width and $CGXO$(X = S/D) is the overlap capacitance (F/m) for the gate-source or gate-drain overlap.

The gate-to-bulk capacitor C_{GB} consists of the series combination of gate-to-channel capacitor C_1 with oxide as the dielectric and the channel-to-bulk capacitor C_2. The gate-to-channel capacitance is given by

$$C_1 = W_{\mathrm{eff}} \cdot (L - 2\mathrm{LD}) \cdot C_{\mathrm{ox}} = W_{\mathrm{eff}} L_{\mathrm{eff}} C_{\mathrm{ox}} \tag{37}$$

Fig. 12 Charge storage capacitances



The channel-to-bulk capacitor C_2 is, however, a depletion capacitor which varies with voltages like C_{DB} and C_{SB}. The capacitance C_2 is given as

$$C_2 = W_{\mathrm{eff}} L_{\mathrm{eff}} \sqrt{\frac{q \varepsilon_{\mathrm{Si}} N_{\mathrm{A}}}{4\Phi_{\mathrm{F}}}} \tag{38}$$

It is important to study the variation of the three intrinsic charge storage capacitances with different regions of operations.

Cutoff: In this region, capacitances C_{GD} and C_{GS} are given by

$$C_{\mathrm{GS}} = C_3 = C_{\mathrm{ox}} W_{\mathrm{eff}} \cdot \mathrm{LD} = CGSO \cdot W_{\mathrm{eff}} \tag{39a}$$

$$C_{\mathrm{GD}} = C_4 = C_{\mathrm{ox}} W_{\mathrm{eff}} \cdot \mathrm{LD} = CGDO \cdot W_{\mathrm{eff}} \tag{39b}$$

$$C_{\mathrm{GB}} = \frac{C_1 \cdot C_2}{C_1 + C_2} \tag{39c}$$

Linear: In the linear region, the channel exists continuously from source to drain and the gate-to-channel capacitance is usually lumped equally into two parts at the source and the drain terminals. Thus

$$C_{\mathrm{GS}} = CGSO \cdot W_{\mathrm{eff}} + 0.5\, C_{\mathrm{ox}} W_{\mathrm{eff}} L_{\mathrm{eff}} \tag{40a}$$

$$C_{\mathrm{GD}} = CGDO \cdot W_{\mathrm{eff}} + 0.5\, C_{\mathrm{ox}} W_{\mathrm{eff}} L_{\mathrm{eff}} \tag{40b}$$

Saturation: In the saturation region, the channel pinches off before reaching the drain and the influence of the drain voltage on either the channel or gate charge is small. As a result, C_{GD} consists of only the overlap component. However, in order to calculate C_{GS}, we compute the total charge Q_{T} stored in the channel. This is given as

$$Q_{\mathrm{T}} = \frac{2}{3} W_{\mathrm{eff}} L_{\mathrm{eff}} C_{\mathrm{ox}} (V_{\mathrm{GS}} - V_{\mathrm{T}}) \tag{41}$$

Therefore, in the saturation region,

$$C_{\mathrm{GS}} = CGSO \cdot W_{\mathrm{eff}} + \frac{2}{3} W_{\mathrm{eff}} L_{\mathrm{eff}} C_{\mathrm{ox}} \tag{42a}$$

$$C_{\mathrm{GD}} = CGDO \cdot W_{\mathrm{eff}} \tag{42b}$$

It is to be noted that the gate-bulk capacitance is neglected in the linear and saturation region. As the gate voltage varies, the charge variation is accounted for by the changes in the charge of the source and the drain regions. The inversion region acts as a shield between the gate and the bulk.

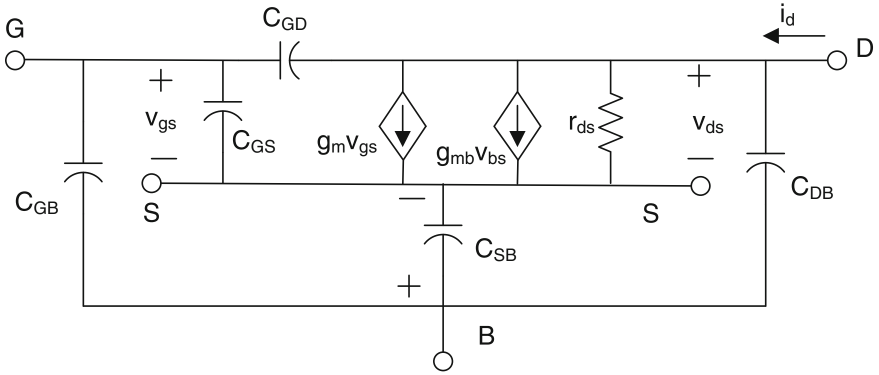


Fig. 13 High-frequency equivalent circuit

The high-frequency equivalent circuit is shown in Fig. 13 including various capacitors. It is, however, quite complex for manual analysis, and is limited to computer-aided simulation through SPICE. This may, however, be simplified, if we assume that the source is connected to the body and neglect the capacitance C_{DB}. Therefore, the capacitances which play significant role in determining the high-frequency response of MOS transistor are C_{GS} and C_{GD}.

5 Switch Model of MOS Transistor

The digital circuit applications of MOS transistor are based on the switching properties of the transistor. Let us consider an n-channel MOS circuit as shown in Fig. 14a. Initially, the transistor is OFF, since $V_{\mathrm{GS}} = 0$. The drain–source voltage of the transistor is initially assumed to be at the supply voltage V_{DD}. The operating

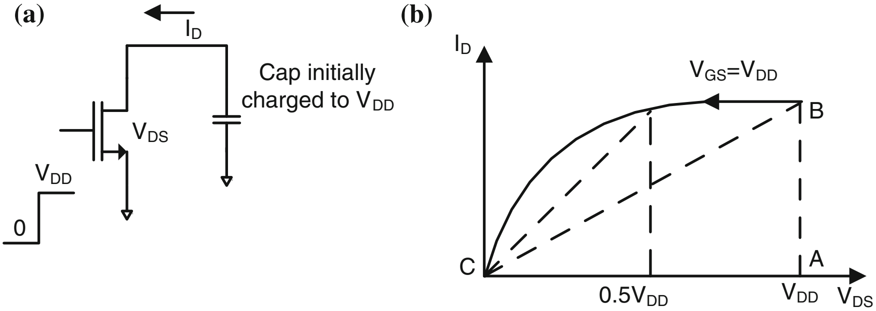


Fig. 14 Switching circuit of MOS transistor, discharging through the transistor: **a** schematic and **b** *I-V* trajectory

point of the transistor under this condition is the point A shown in Fig. 14b. If the gate voltage is switched from 0 toV_{DD}, the initial drain current is given by

$$I_{\mathrm{D}} = \frac{\beta_{\mathrm{n}}}{2}(V_{\mathrm{DD}} - V_{\mathrm{Tn}})^2 \tag{43}$$

After switching takes place, the MOS transistor becomes ON, the operating point moves from the point A to the point B. The capacitor discharges through the transistor and thus the operating point follows the curve $V_{\mathrm{GS}} = V_{\mathrm{DD}}$ down to the point C, where $I_{\mathrm{D}} = 0$, $V_{\mathrm{DS}} = 0$. The drain–source switching resistance is time varying, nonlinear, and dependent on the operating point of the transistor. A logical approach for determining the resistance is therefore to find out the average value of the resistance at the end points of the transitions assuming that it does not experience any strong nonlinearities over the range of the averaging interval. This is given as the reciprocal slope of the line BC in Fig. 14b

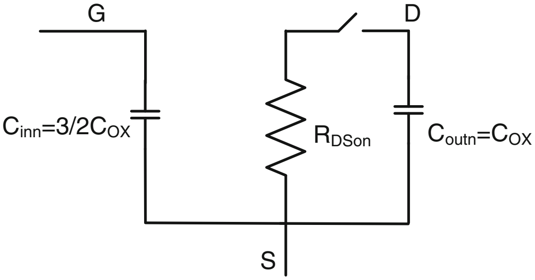
$$R_{\mathrm{DSon}} = \frac{2V_{\mathrm{DD}}}{\beta_{\mathrm{n}}(V_{\mathrm{DD}} - V_{\mathrm{Tn}})^2} = R'_{\mathrm{DSon}}\frac{L}{W} \tag{44}$$

We observe from (44) that the drain–source switching resistance is inversely proportional to the W/L ratio of the transistor. In addition, as the supply voltage approaches the threshold voltage, the resistance increases dramatically.

In order to estimate the capacitances of the switching MOS transistor, let us consider the transistor with gate–source capacitor C_{GS} and gate–drain capacitor C_{GD}. From our previous discussion, we see that in the linear region, the values of both these capacitors are considered to be $C_{\mathrm{OX}}/2$, where $C_{\mathrm{OX}} = C_{\mathrm{ox}} \cdot W \cdot L$. Since this overestimates the overlapping capacitance, we neglect the depletion capacitances that exist between the source and drain regions with the substrate. The gate–drain capacitor C_{GD} is the Miller capacitor and the corresponding impedance is ($Z = 1/sC_{\mathrm{GD}}$). The large-scale gain (K) between the input and the output of the transistor, when it acts as a switch is always equal to −1. Thus according to Miller's theorem, the impedance K is broken into two impedances Z_1 and Z_2 between the gate to ground and between the drain to ground, respectively. The values of these impedances are given as $Z_1 = Z/(1 - K)$ and $Z_2 = Z/(1 - 1/K)$. In other words, the capacitance C_{GD} is splitted into two capacitances of values $2C_{\mathrm{GD}}$ or C_{OX}; one from gate to source and the other from the drain to source. Thus the input capacitance value becomes $C_{\mathrm{inn}} = C_{\mathrm{GS}} + 2C_{\mathrm{GD}} = 3/2C_{\mathrm{OX}} = 3/2C_{\mathrm{ox}} \cdot W \cdot L$ and that of the output capacitance becomes $C_{\mathrm{outn}} = 2C_{\mathrm{GD}} = C_{\mathrm{OX}} = C_{\mathrm{ox}} \cdot W \cdot L$.

The complete model of a switching MOS transistor is shown in Fig. 15. When $V_{\mathrm{GS}} > \frac{1}{2}V_{\mathrm{DD}}$, the switch is closed and when $V_{\mathrm{GS}} < \frac{1}{2}V_{\mathrm{DD}}$, the switch is open. It is to be noted while deriving this model, the rise time of the input transition is assumed to be zero. Although, this does not occur in reality, the model provides remarkably good results compared to simulation or measurement results.

Fig. 15 Simple switch model of an MOS transistor



5.1 Intrinsic Switching Delay and Speed of MOS Transistor

Let us consider that the drain of an n-channel MOS transistor is charged to V_{DD} and the input gate terminal switches from 0 to V_{DD}. It follows from the simple switch model that the output drain voltage will decay through the drain–source switching resistance with a time constant $R_{\mathrm{DSon}}C_{\mathrm{OX}}$. For an n-channel transistor, the intrinsic switching delay is given as

$$\tau_{\mathrm{n}} = R_{\mathrm{DSon}}C_{\mathrm{OX}} = \frac{2L^2}{\mu_{\mathrm{n}}V_{\mathrm{DD}}\left(1 - \frac{V_{\mathrm{Tn}}}{V_{\mathrm{DD}}}\right)^2} \tag{45}$$

The intrinsic speed is proportional to the reciprocal of τ_{n}. We observe from (45) that the intrinsic speed of an MOS transistor is independent of the channel width, increases inversely as the square of the channel length. The intrinsic speed of an n-channel transistor is higher than that of a p-channel transistor due to enhanced mobility of the former compared to the later. It is seen that the delay increases with supply voltage in a nonlinear manner and it increases sharply as the supply voltage approaches the threshold voltage.

5.2 Switching Parameters of MOS Transistor

The three important switching parameters of an MOS transistor are delay and rise time, fall time. In the following sub-subsection, the measurement procedures of these parameters are first defined.

5.2.1 Definition of Delay and Rise Time, Falltime

The definitions that are provided here are valid for all types of logic circuits. Consider Fig. 16. The upper trace represents the input to any logic gate while the lower one represents the output. The input rise and fall times are denoted by t_{r} and

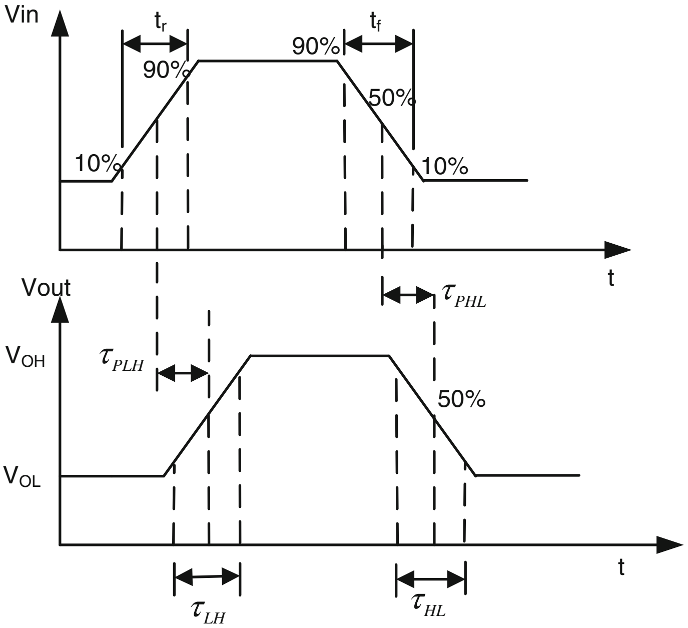


Fig. 16 Definition of delay, input rise and fall times, and output rise and fall times

$t_{f,}$ respectively. Similarly, the output rise and fall times are denoted by τ_{LH} and $\tau_{\mathrm{HL},}$ respectively. The uncertainty over when the transition actually starts or ends is avoided by defining the rise and fall times between the 10 and 90 % points of the waveforms. The propagation delay τ_{P} of a gate defines how quickly it responds to a change at its input(s). This quantifies the delay experienced by a signal when passing through a gate. The delay time between the 50 % points of the input and the output waveforms are denoted by τ_{PLH} and τ_{PHL} depending on whether the output is changing from low to high or from high to low. The propagation delay τ_P is defined as the average of the two: $\tau_{\mathrm{P}} = (\tau_{\mathrm{PLH}} + \tau_{\mathrm{PHL}})/2$. It is to be noted that the propagation delay parameter τ_{P} in contrast toτ_{PLH} and τ_{PHL} has no physical significance as such. It is mostly used to compare different semiconductor technologies, circuit, or logic design styles. The propagation delay of a gate depends upon circuit topology and technology. In addition, it depends upon the rise time of the input and output signals of a gate.

5.2.2 Determination of Switching Parameters for an MOS Transistor

Digital gates are often modeled as a first-order RC network as shown in Fig. 17.

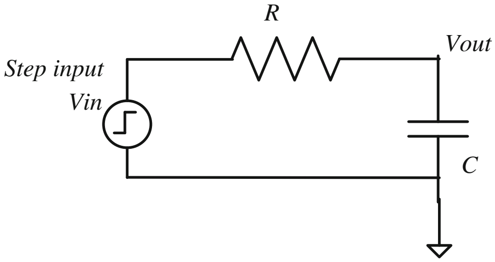


Fig. 17 Delay and rise time for a first-order RC network

With the application of a step input voltage (V_{in} switching from 0 to V), assuming that the capacitor is initially uncharged, the output voltage across the capacitor is given as

$$V_{\text{out}} = V\left(1 - \mathrm{e}^{-\frac{t}{\text{CR}}}\right) \tag{46}$$

For this simple circuit, the delay time, i.e., the time to reach the 50 % point is easily computed as $t = \ln(2)\text{RC} = 0.69$ RC. Similarly, the rise or fall time is given by $t = \ln(9)\text{RC} = 2.2$ RC to get to the 90 % point.

For the simple switching model of a transistor as shown in Fig. 15, following similar analysis as presented above, the propagation delay time, whether high to low or low to high is given as

$$\tau_{\text{PHL}}, \tau_{\text{PLH}} = R_{\text{DSon}} \cdot C_{\text{tot}} \tag{47}$$

The output rise and fall times are given by

$$\tau_{\text{HL}}, \tau_{\text{LH}} = 2R_{\text{DSon}} \cdot C_{\text{tot}} \tag{48}$$

In (47) and (48), C_{tot} is the total capacitance from the drain of the MOS transistor to ground. Depending on cases, R_{DSon} is for n-channel or p-channel MOS transistor. These equations for calculating the delay and rise time and fall time are however, approximate ones that are found to give good first-hand results quickly.

5.3 Switching Strength of MOS Transistors

The strength of a signal is measured by the degree of closeness of the magnitude of the signal with the magnitude of an ideal voltage source. The power supplies, or rails, (V_{DD} and GND) are the source of the strongest "1"s and "0"s.

An n-channel MOS transistor behaves as a perfect switch when a logic "0" signal is passed through it and thus it is said to pass a strong "0". On the other hand, it behaves imperfectly when a logic "1" signal is passed through it and thus it is said to pass a weak "1". It may be noted that an n-channel MOS transistor turns ON when a

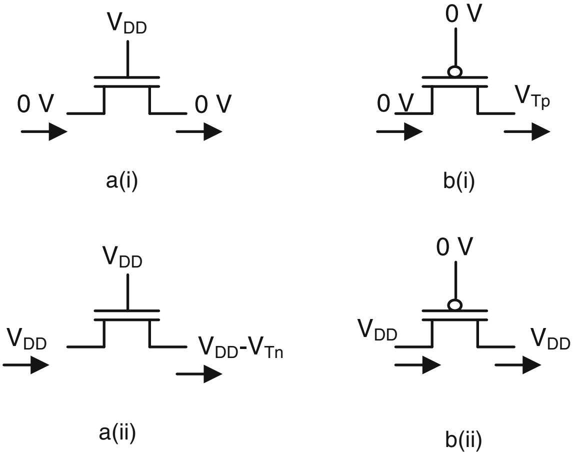


Fig. 18 Switching strengths of MOS transistors: **a(i)–a(ii)** strong "0" and weak "1" of n-channel MOS transistors; **b (i)–b(ii)** weak "0" and strong "1" of p-channel MOS transistors

logic "1" signal is applied at the gate terminal. This is illustrated in Fig. 18a(i) and a (ii). It is observed that logic "0" is perfectly passed because the n-channel MOS transistor remains ON always. On the other hand, while passing logic "1" signal, the output cannot go above $V_{DD} - V_{Tn}$, in order to keep the transistor ON.

A p-channel MOS transistor exhibits the reverse behavior, passing strong "1"s but weak "0"s. It may be noted that a p-channel MOS transistor turns ON when a logic "0" signal is applied at the gate terminal. This is illustrated in Fig. 18b(i) and b (ii). It is observed that when a logic "0" signal is to be passed, the output cannot go below V_{Tp} in order to keep the p-channel MOS transistor ON. On the other hand, while passing logic "1" signal, the signal is fully passed at the output, because the p-channel MOS transistor always remain ON.

It is because of this behavior, n-channel MOS and p-channel MOS transistors are operated in a complimentary manner in digital circuits. When any node of a circuit needs to be pulled up to the power supply voltage, it is connected through a p-channel MOS transistor. On the other hand, if it needs to be pulled down to the ground voltage, it is connected through an n-channel MOS transistor. Thus p-channel MOS transistors are used to implement the pull-up network of a digital circuit and n-channel MOS transistors are used to implement the pull-down network.

6 Compact Device Modeling

As emphasized already, the MOS transistor acts as the basic building block for VLSI circuit simulations. For circuit simulation purpose, the actual device physics of MOS transistor is abstracted in the form of device models. Device models are considered as the link between the physical world (technology, manufacturing etc.)

and the design world (device simulation, timing simulation, circuit simulation) of the semiconductor industry. The fundamental physical characteristics of MOS transistors are described by a set of mathematical equations producing a compact device model for circuit simulation. The compact model should be able to faithfully represent the device characteristics for different device geometries, range of temperature and process variations etc. The ability to create an accurate model of a device is just as critical as the ability to manufacture it.

6.1 Categories of Compact Models

The compact models of MOS transistors which are available in the public domain are broadly classified into three categories: (i) threshold voltage V_T-based models, (ii) charge, Q_n-based models, and (iii) surface potential, ψ_s-based models. The threshold voltage-based model is a piece-wise approach for describing the transistor behavior in the sense that the weak and strong inversion region is demarcated by threshold voltage. Berkeley's Level 1, Level 2, Level 3 BSIM1, BSIM2, BSIM3, and BSIM4 belongs to the category of threshold voltage-based compact model. This category of compact model although suffers from the limitation of having large number of fitting parameters, yet remains the de facto industry standard until very recently. In the charge-based compact model, the drain current is formulated in terms of the inversion channel charge density at the source and drain ends. The EKV compact model belongs to the category of the charge-based compact model. The latest version of BSIM, i.e., BSIM6 belongs to this category. This category of compact model has the advantage that it is physics based with a minimum requirement of empirical fitting. On the other hand, in the surface potential-based model, the drain current is formulated in terms of the surface potential at the source and drain ends of the channel. The PSP and HiSIM models belong to the category of surface potential-based compact model.

6.2 Use of Compact Model in SPICE Circuit Simulation

Circuit simulation is an essential process in designing integrated circuits. The use of a compact device model in VLSI circuit simulation is illustrated in Fig. 19. The simulation process starts with a set of desired specifications for the circuit to be designed. Depending upon it, a certain topology of the circuit is selected by the designer, which is then described through SPICE netlist. The SPICE circuit simulator is used to simulate the circuit. The SPICE simulation tool internally uses a set of compact device models for the components of the circuit and solves a set of network theory equations through standard numerical algorithms.

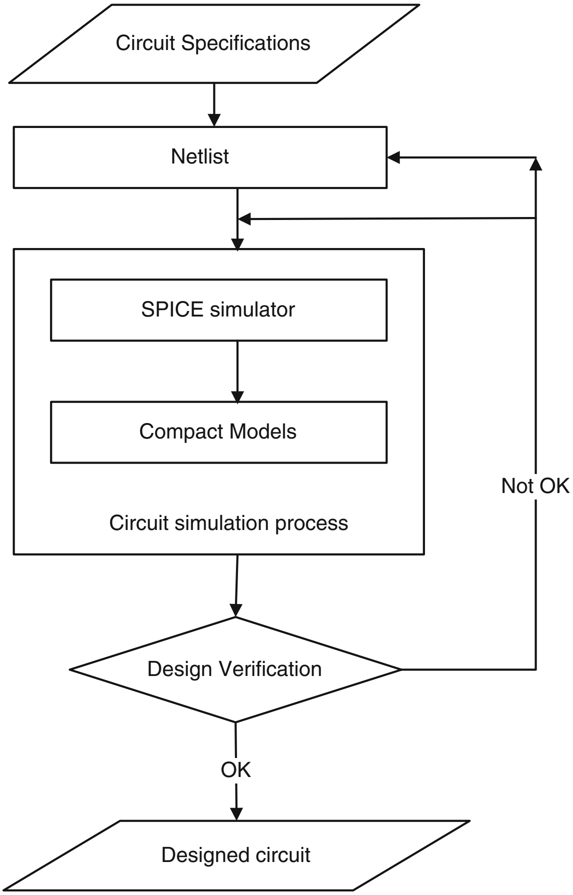


Fig. 19 Use of compact device model in circuit simulation through SPICE

If the designer is satisfied with the performance of the selected topology against the desired specifications, the simulation process is stopped; otherwise, the process is iterated. The performance of the circuit is heavily dependent on the accuracy of the chosen device compact model. This illustrates the use of compact device models in circuit simulation process.

6.3 Model Standardization

Several models are available in the literature, explaining qualitatively the different phenomena involved with an MOS transistor. Since 1996, an independent Compact Model Council, consisting of several different semiconductor industries and research organizations has spearheaded the model standardization process. The members and consultants involved with CMC have collaborated to develop and standardize compact models for widely used semiconductor devices. In 2013, the Silicon Integration Initiative (Si2) acquiesced the Compact Model Council. It is renamed as the Compact Model Coalition. The CMC released the BSIM6 model to the industry and is now in the process of standardizing models for extremely thin SOI and GaN FET.

Further Reading

1. Y. Taur, T.H. Ning, *Fundamentals of Modern VLSI Devices*, 2nd edn. (Cambridge University Press, Cambridge, 2013)
2. A.B. Bhattacharya, *Compact MOSFET Models for VLSI Design* (Wiley, New York, 2009)
3. Y. Tsividis, C. McAndrew, *Operation and Modeling of the MOS Transistor*, 3rd edn. (Oxford University Press, Oxford, 2010)
4. S. Pandit, C. Mandal, A. Patra, *Nano-Scale CMOS Analog Circuits Models and CAD Techniques for High-Level Design* (CRC Press, USA, 2014)
5. R. Jacob Baker, *CMOS Circuit Design, Layout and Simulation*, 3rd edn. (Wiley-IEEE Press, New York, 2011)
6. P.E. Allen, D.R. Holberg, *CMOS Analog Circuit Design*, 3rd edn. (Oxford University Press, Oxford, 2011)
7. C.K. Sarkar, *Technology Computer Aided Design Simulation for VLSI MOSFET* (CRC Press, USA, 2013)
8. N.Weste, D.Harris, *CMOS VLSI Design; A Circuits and Systems Perspective*, 4th edn. (Pearson Education, 2011)

Applications of Nanotechnology in Next-Generation Nonvolatile Memories

Amretashis Sengupta, Bikash Sharma and Chandan Kumar Sarkar

1 Introduction

Keeping record of objects, events, and the knowledge gathered from their observation has long been the habit of the human species. In this continued effort of data storage, mankind has made a remarkable journey from cave paintings and clay tablets in prehistory, to the printing press in the industrial age, and in this modern electronic age to the flash memory devices. As technology has evolved and new fabrication techniques and new materials have come into play, MOS devices which are the cornerstone of the electronic age, have become smaller and smaller.

This continued miniaturization or "scaling" in the VLSI industry [1, 2] has greatly impacted the cost and performance of the MOS devices in general, and also that of MOS nonvolatile memory (NVM) cells. This has been a catalyst in making MOS NVM, the most widely used storage technology in modern consumer electronic gadgets. The variety of memory cards like the SD, micro SD, MMC, M2, and the ubiquitous USB flash drive are quite cheap, very convenient to store and carry, and reliable enough to retain data fidelity for a very long time. Thus, MOS NVM devices have become the storage media of choice for everyday operations.

Among the various kinds of architectures of MOS NVM cells, the floating gate and the SONOS NVM cells have been the most widely used in the industry [3–8].

A. Sengupta (✉)
School of VLSI Technology, Indian Institute of Engineering Science and Technology,
Shibpur, Howrah 711103, India
e-mail: a.sengupta@vlsi.iiests.ac.in

B. Sharma · C.K. Sarkar
Department of Electronics and Telecommunication Engineering, Jadavpur University,
Kolkata 700032, India
e-mail: phyhod@yahoo.co.in

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_8

However, as device feature sizes get smaller and smaller, the so-called "scaling problems" like leakage through ultrathin tunnel dielectrics and reliability issues such as lowering of time-to-breakdown arise [6, 8]. These reasons have prompted different approaches toward making next-generation NVM cells more immune to scaling challenges. The embedding of semiconducting/metallic nanocrystals in the gate dielectric of MOS NVM cells has yielded some success in addressing the problems of leakage, lateral diffusion, and breakdown [9–40]. However, the size of the nanocrystals remains a hindrance in further scaling of these devices. With the introduction of multigate architectures in MOS NVM cells and the advent of tunnel FET memory devices, the reign of MOS memory could be extended by a few more years [41–53].

As scientists and engineers look to sustain Moore's law beyond the 22 nm technology node, memory devices based on principles rather different from charge storage to represent data (as in MOS NVM devices) are coming to the fore [8]. Resistive RAM (RRAM) technology employing MIM capacitors of materials having distinguishable high-resistance states (HRS) and low-resistance states (LRS) recently has emerged as an alternative NVM with simple structure, high speeds, and good CMOS compatibility [54–69]. Also by switching the polarization state of the Ferroelectric materials such as $BaTiO_3$, $LiNb_2O_5$, PZT ($PbZr_xTi_{1-x}O_3$), BST ($Ba_{1-x}Sr_xTiO_3$), $(Pb,La)(Zr,Ti)O_3$ or PLZT, ploy(vinylidene fluoride trifluoroeithylene), or P(VDF-TrFe), the Ferroelectric FET (FeFET) memories have been developed [70–82]. Moreover, the switching between low-resistive parallel and high resistive antiparallel states in magnetic tunnel junction (MTJ) by the mechanism of spin transfer torque (STT) has revived magnetic memories in a big way [83–102].

The CNT FET or nanowire FET, offer certain advantages over the present advanced MOSFETs like the DGMOSFET/GAA MOSFET or the FinFET. Due to the almost defect-free nature of CNT/graphene, the electron transport in such structures fall in the ballistic transport regime, thereby enabling very high speed. With modern ULSI technology approaching the 22 nm node, the CNT- and graphene-based nanotransistors may well be able to operate at gate lengths of 10 nm. In this context, CNT and graphene-based memories promise a superior performance than their conventional Si counterparts [103–110].

More recently, nanoelectromechanical memories based on the principle of charge trapping and detrapping based on mechanical contact and nanomanipulation of nanorods/nanoparticles encapsulated in coaxial CNT structures are undergoing research and development [111–116].

These various types of NVM devices rely on different properties and mechanisms to store information, and with so many types of emerging NVMs, it is necessary to understand the structure, operating principles, and the pros and cons of each type in a concise way. This chapter attempts to keep the reader abreast with the latest in the NVM sector by providing a concise and easy to understand overview of these modern memory devices from the point of their architecture, fabrication methods, the principle of operation, and its underlying physics.

2 Nanocrystal-Embedded MOS NVM

The gate dielectric in a conventional SONOS/Floating-gate MOS consists of successive layers of tunnel oxide-nitride or polysilicon floating gate-control oxide and metal or polysilicon contact [8]. In such devices, depending upon the substrate type (p-Si or n-Si) upon the application of a sufficient gate bias electrons (or holes), quantum mechanically tunnel across the tunnel oxide layer and get trapped in the potential well created by the floating gate. This stored charge induces a shift in the flatband voltage {*i.e., the gate voltage required to induce zero net charge in the underlying semiconductor in a MOS capacitor*} and the threshold voltage {*i.e., the minimum gate voltage required to form an inversion layer at the interface between the gate dielectric and the substrate in a MOSFET*}. And therefore, the charged and discharged states of the MOS device can be distinguished to represent data bits in an NVM cell (Fig. 1).

As device dimensions get smaller and smaller, several problems arise in these conventional FG MOS devices, with the ultrathin oxide layers in the ULSI technology. One of them being the leakage of stored charge through this ultra-thin tunnel dielectric. Also there exists the problem of lateral diffusion leakage in non-isolated MOS NVM structures. The repeated Program/Erase cycles, in conventional planar SONOS/FG-MOS generates a high amount of electrical stress on tunnel dielectrics and gives rise to serious reliability issues such as smaller time to breakdown.

Thus a solution is proposed by embedding nanoparticles of semiconducting or metallic materials in a thin laminar layer of the gate oxide (usually SiO_2 or in some cases High-k materials like HfO_2) to effectively serve as a floating gate (Fig. 2).

In this way, the charge storage nodes can be made discrete rather than a single continuous layer for better breakdown resistance and minimization of lateral diffusion. Also such devices show better charge retention by breaking the continuous tunneling path from the substrate to the gate. Tiwari et al. [9] were the first to successfully fabricate a nc-Si-embedded SiO_2 gate oxide Silicon MOS NVM in 1996. Since then, academic and industrial interest in such nanocrystal-embedded MOS memory has grown significantly. Major industry players like Samsung,

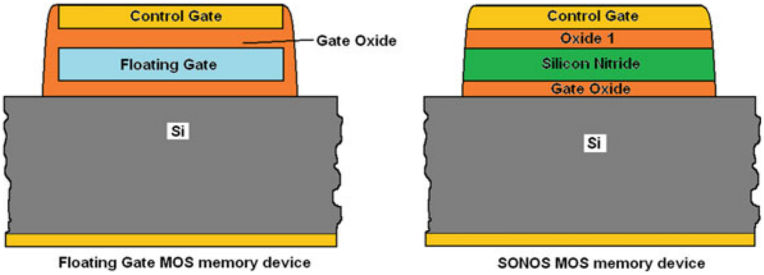


Fig. 1 SONOS and floating gate MOS memory devices

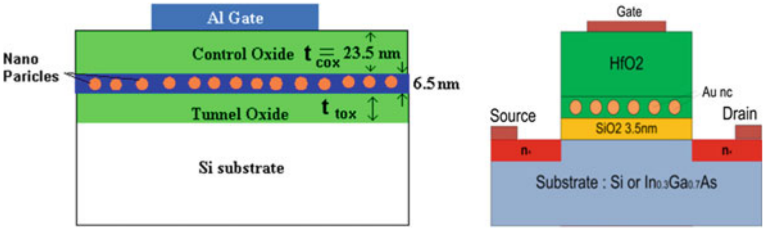


Fig. 2 Nanocrystal-embedded gate dielectric MOSCAP and MOSFET NVMs

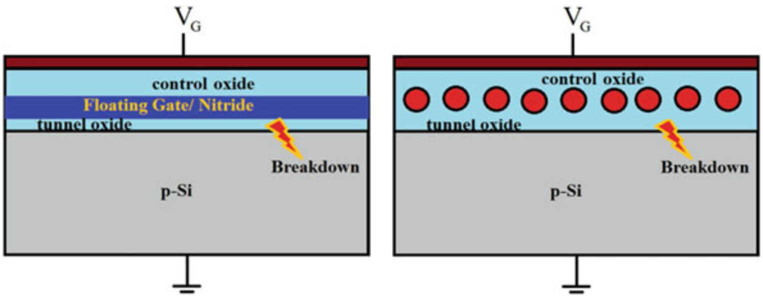


Fig. 3 Advantages in breakdown resistance and minimization of lateral diffusion in nc-embedded MOS NVMs: local breakdowns in tunnel oxide compromise the entire FG/nitride layer in conventional cells, but only a few nanocrystals in nc-embedded devices

Motorola, Freescale, and IBM have been actively involved in the development of this class of MOS NVM, with Freescale bringing the first nc-Si-embedded MOS NVM into the market in 2007 (Fig. 3).

In addition to the advantage of lesser leakage, improved reliability, and suppression of lateral diffusion, there is the option of further improvement with multilayer embedding of nanocrystals to create multiple charge trap wells through work function engineering. Also by varying the areal density and size distribution of embedded ncs, it is possible to fabricate MOS NVMs of different capacitance values, for identical gate dielectric thickness (Fig. 4).

2.1 Write/Erase Mechanism of nc-Embedded MOS NVMs

To explain the write (program) and erase mechanism of the nc-embedded MOS NVM, we consider a p-Si MOS NVM with nanocrystals embedded in the gate dielectric. In such p-Si MOS devices, under the application of a positive gate bias, a triangular potential well is created on Si side of the Si-tunnel oxide interface. The energy states of the so-formed inversion layer are simply the Airy functions

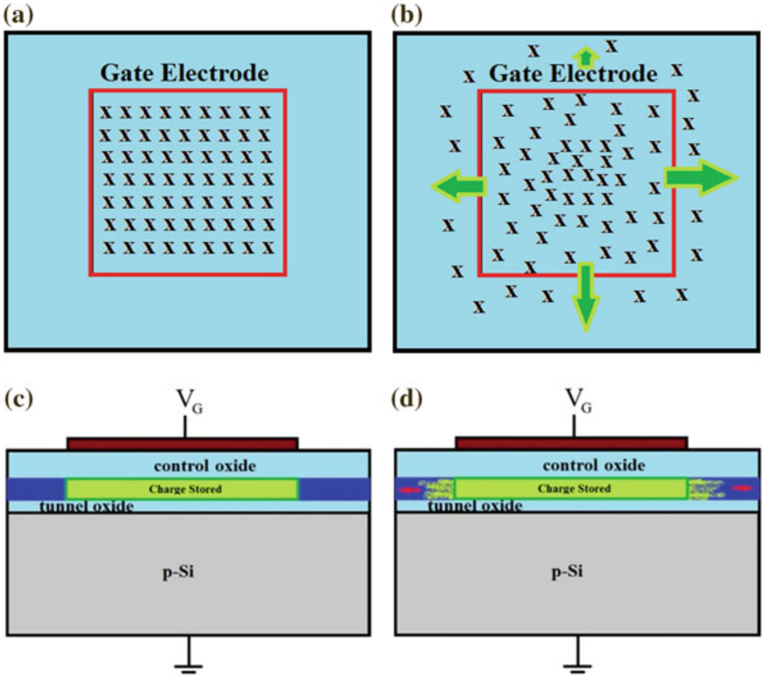


Fig. 4 Schematic top view and side view of the lateral diffusion mechanism (**a**, **c**) after writing $t = 0$ the charge is concentrated under the gate area but with (**b**, **d**) time elapsed $t > 0$ the charge diffuse away from area directly under the gate electrode

obtained by solving the Schrödinger equation for a triangular potential well [116–121]. These levels are expressed as $E_{n,\ell}$ which is the energy of the electron in the ℓth valley of the nth level $(n = 1, 2, 3\ldots)$

$$E_{n,\ell} = \left(\frac{\hbar^2}{2 \cdot m_{X\ell}}\right)^{\frac{1}{3}} \cdot \left(\frac{3}{2}\pi q F_{\mathrm{Si}} \cdot \left(n - \frac{1}{4}\right)\right)^{\frac{2}{3}} \tag{1}$$

Here q is the electronic charge, F_{Si} is the electric field in the Si substrate, $m_{X-\ell}$ is the effective mass of the electron in the ℓth valley in the direction perpendicular to the Si–SiO_2 interface. It is the electron tunneling from this well into the nanocrystal energy levels (through the tunnel oxide) that determines the charging of the MOS NVM.

In such devices, depending upon the substrate type (p-Si or n-Si) upon the application of a sufficient gate bias electrons (or holes) quantum mechanically tunnel across the tunnel oxide layer and get trapped in the potential well created by the floating gate. This stored charge induces a shift in the flatband voltage {*i.e., the gate voltage required to induce zero net charge in the underlying semiconductor in a MOS capacitor*} and the threshold voltage {*i.e., the minimum gate voltage* *required to form an inversion layer at the interface between the gate dielectric and the substrate in a MOSFET*}. And therefore, the charged and discharged states of the MOS device can be distinguished to represent data bits in an NVM cell (Fig. 5).

Depending upon the device structure, substrate, oxide properties, and the nc chosen, the tunneling barriers vary in depth and thickness, and therefore the tunneling currents are also affected. The basic principle of operation of such devices however remains the same, as described above.

Fowler Nordheim Tunneling: In order to express the electron tunneling for a high applied field across the gate dielectric, as in the case of MOS NVMs, where several volts are applied as write voltage across few tens of nanometers thick gate dielectric, the Fowler–Nordheim tunneling (F–N tunneling) mechanism has to be considered. According to the works of Lenzlinger and Snow [119] the F–N mechanisms in MOS devices occur usually at high applied electric fields ($\sim 10^6$ V/cm). At such fields, the electrons face a triangular barrier, as shown in Fig. 6.

In this type of tunneling, as the electrons tunnel through a triangular barrier, they tunnel across a lesser distance than the actual layer thickness. Also, there exists a field-induced barrier lowering in case of F–N tunneling, which further promotes the barrier transmission.

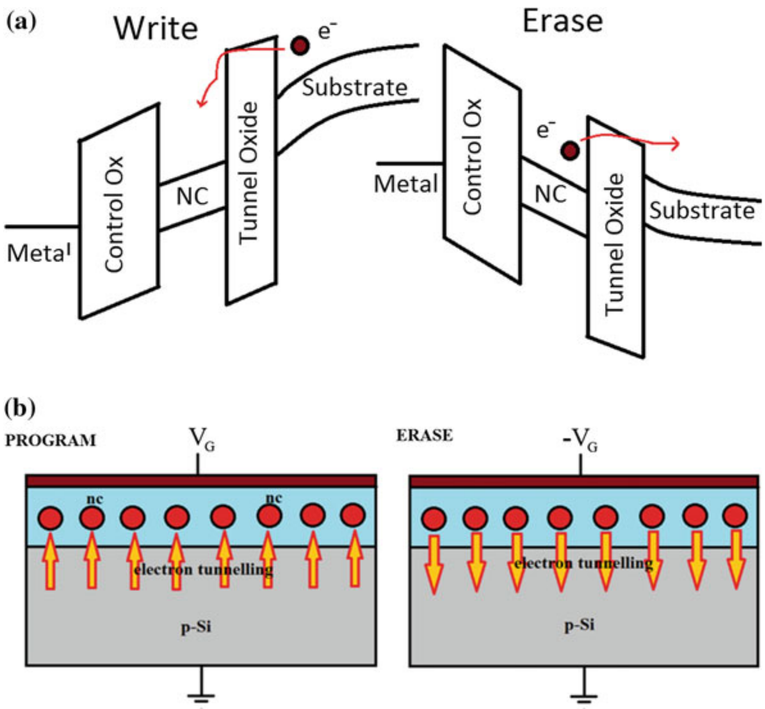
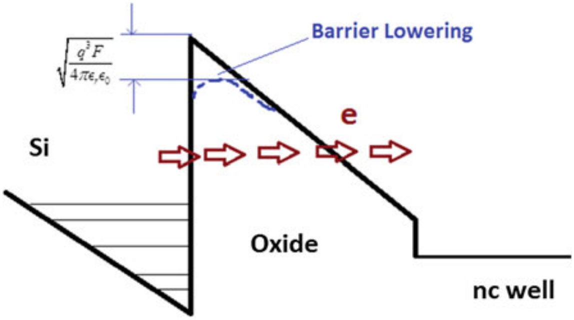


Fig. 5 Write (program) and erase mechanisms of a nc-embedded p-MOS NVM. **a** Energy band diagrams. **b** Pictorial representation using the device schematic for a nc-embedded MOSCAP

Fig. 6 Typical energy band diagram of electron tunneling by F–N mechanism in a p-Si MOS structure



For getting the effective electron tunneling barrier height φ'_{eff}, we have to consider the image force term and also the barrier lowering induced by the F–N mechanism.

$$\varphi'_{\text{eff}} = \varphi_{\text{eff}} - \beta F_{\text{eff}}^{1/2} - \alpha F_{\text{eff}}^{2/3} \tag{2}$$

Here, φ_{eff} is the intrinsic oxide-Si barrier height, F_{eff} is the oxide effective electric field whereas α and β are constants. The second term in the above equation arises out of the image force, and the third term comes due to the Fowler–Nordheim barrier lowering mechanism. The value of β is calculated from the following expression, [119–121]

$$\beta = \left(\frac{q^3}{4\pi k_{\text{eff}} \in_0}\right)^{1/2} \tag{3}$$

$\in_0$ is the free space dielectric constant, β is expressed in $\text{e}(\text{V cm})^{1/2}$. However, for α there is no such simple expression, hence the experimentally fitted values of α for conventional MOS structure ($\alpha \approx 1.0 \times 10^{-5}\ \text{e}(\text{V cm}^2)^{1/3}$) have to be used [117–124].

The general expression for F–N tunneling through the triangular energy barrier is obtained from the Wentzel–Kramers–Brillouin (WKB) approximation as

$$D(E_{n,\ell}) = \exp\left\{-\frac{2}{\hbar}\int_{x1}^{x2} \sqrt{2m\left[V(x) - E_{n,\ell}\right]} \cdot \mathrm{d}x\right\} \tag{4}$$

where m is the effective electron mass in the region between x_1 and x_2, $V(x)$ is the potential. $E_{n,\ell}$ is the energy of the electron in the ℓth valley of the nth level ($n = 1, 2, 3\ldots$) in the well created in the silicon inversion layer. The expression for F–N tunneling current density thus comes out to be [118–123]

$$J_{\mathrm{F-N}} = \sum_{n,\ell} \left(\frac{q^3 m_{\mathrm{si}}}{16\pi^2 m_{\mathrm{ox}} \hbar} \right) \cdot \frac{F_{\mathrm{ox}}^2}{\left(\varphi_{\mathrm{eff}} - E_{n,\ell}\right)} \cdot \exp\left\{ -\left(\frac{2\sqrt{8 m_{\mathrm{ox}} q}}{3\hbar} \right) \frac{\left(\varphi_{\mathrm{eff}} - E_{n,\ell}\right)^{3/2}}{F_{\mathrm{ox}}} \right\}, \tag{5}$$

where F_{ox} is the effective electric field across the oxide, m_{si} and m_{ox} are the electronic effective masses in the p-Si substrate and the gate dielectric, $\hbar$ is the reduced Planck's constant, and q is the electronic charge.

For the erase operation, a sufficient negative gate bias is applied to the p-Si MOS NVM, which causes F–N tunneling of electrons trapped in the nc energy states back into the Si substrate. The erase current density is evaluated as

$$J_{\mathrm{erase}} = \left(\frac{q^3 m_{\mathrm{nc}}}{16\pi^2 m_{\mathrm{tox}} \hbar} \right) \cdot \frac{F_{\mathrm{tox}}^2}{\left(\varphi_{e-\mathrm{tox}} - \varepsilon_{\mathrm{nc}}\right)} \cdot \exp\left\{ -\left(\frac{2\sqrt{8 m_{\mathrm{tox}} q}}{3\hbar} \right) \frac{\left(\varphi_{e-\mathrm{tox}} - \varepsilon_{\mathrm{nc}}\right)^{3/2}}{F_{\mathrm{tox}}} \right\}. \tag{6}$$

In (6) $\phi_{e-\mathrm{tox}}$ is the barrier height for electron tunneling from the nc energy state back into the Si substrate $\varepsilon_{\mathrm{nc}}$ is the nanocrystal energy states, m_{tox} is the electron effective mass in the tunnel oxide, and F_{tox} is the field across the tunnel oxide.

Another point worth mentioning here is that the application of F–N tunneling to write/erase data in these type of devices results in a high F–N stress on the tunnel oxide structure, and therefore the write voltage/erase voltage is selected as the voltage at which the F–N currents just starts to dominate the DT current [26, 27, 31]. This is basically the "onset" voltage for F–N tunneling mechanism. Usually, for such nc-embedded gate dielectric MOS NVM devices, this voltage is usually associated with an electric field of $\sim 5 \times 10^6$ V/cm [3–8].

Direct Tunneling: When the write operation is complete, the nc states have been charged by the electrons from the Si inversion layer well by the F–N tunneling mechanism. As the device is in the OFF state (data-retention state) and no significant gate bias has been applied, this stored charge in the nc states tends to "leak" back into the Si substrate tunneling through the trapezoidal shaped barrier by the so-called "Direct tunnelling" mechanism. This is a kind of low-field tunneling that occurs as a result of the low electric field created in the nc to Si direction by the charged nc-embedded layer (Fig. 7).

The expression for such direct tunneling current is given by

$$J_{\mathrm{DT}} = \frac{\left\{2 m_{\mathrm{ox}} \left(\phi_{\mathrm{eff}} - \varepsilon_{i,k}\right)\right\}^{1/2} \alpha q^2 V}{\hbar^2 d} \exp\left(\frac{2\alpha \sqrt{2 m_{\mathrm{ox}} \left(\phi_{\mathrm{eff}} - \varepsilon_{i,k}\right)}}{\hbar} \cdot d \right), \tag{7}$$

where d is the tunnel oxide thickness, V is the applied bias, $\varepsilon_{i,k}$ is the nanocrystal energy states, and α is a dimensionless adjustable parameter introduced to account for the effective mass [121].

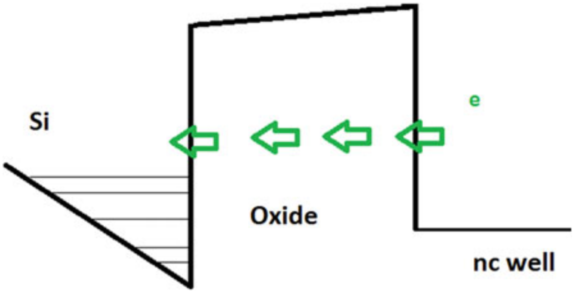


Fig. 7 Schematic of the direct tunneling (DT) mechanism

Now, the application of such a write voltage causes, the electrons from the substrate inversion layer, to tunnel across the tunnel oxide and be captured by the nc states. The Flatband voltage shift caused by the accumulation of charge in the nc-embedded layer is given by [19–21, 25, 28, 31]

$$\delta V_{\mathrm{FB}} = \frac{Q_{\mathrm{NC}}}{\in_{\mathrm{ox}}}\left(t_{\mathrm{cox}} + \frac{1}{2}\cdot\frac{\in_{\mathrm{ox}}}{\in_{\mathrm{nc}}}\cdot D\right) \qquad (8)$$

In Eq. (8), Q_{NC} is the charge stored in the nc states, $\in_{\mathrm{nc}}, \in_{\mathrm{ox}}$ are the dielectric constants of the nc, and the oxide, respectively, D is the average diameter of the nanocrystals, t_{cox} being the thickness of the control oxide. In case of MOSFET memories, however, it is the threshold voltage shift which determines the state of the memory (Fig. 8).

Another parameter of importance in this regard is the memory window which is defined as the change in the threshold/flatband voltage of the device as the gate voltage swings from positive (electron charging) to negative polarities (hole charging) of V_{dd} (Fig. 9).

A higher amount of flatband/threshold voltage shift and a temporally stable memory window indicate a good MOS NVM device.

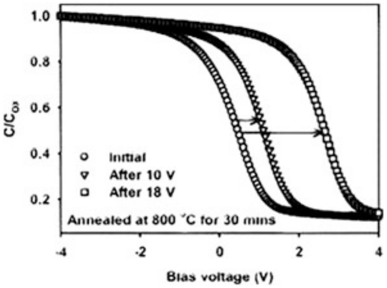


Fig. 8 Flatband voltage shifts due to charging of ncs (reprinted with permission from Yang et al. [16]). © AIP 2009

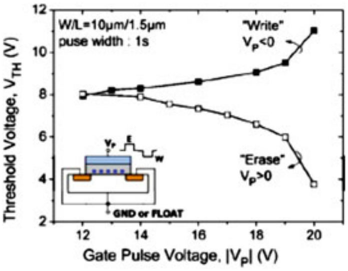


Fig. 9 Write/erase characteristics of MOSFET NVM (reprinted with permission from Koliopoulou et al. [32]). © Elsevier 2006

2.2 *Standard Fabrication Techniques for nc-Embedded MOS NVMs*

These nc-embedded gate dielectric layers in such MOS devices are usually fabricated by methods like ion implantation, RTA size-selective precipitation of nanocrystals, or molecular self-assembly by template growth. The typical diameter of ncs used in such devices is in the range of 3–6 nm. The total gate dielectric thickness is usually between 25 and 45 nm [9–23].

This is one of the most widely used processes used to form a layer of nanodots embedded within a host matrix of SiO_2 or some other gate dielectric for NVM applications. In this process, first the tunnel oxide layer is thermally grown on the Si substrate. The material for the nanocrystal (say Au) is then deposited in the form of an ultrathin film by ALD or sputtering or CVD on top of the tunnel oxide. This ultrathin film is then subjected to rapid thermal annealing (RTA) in temperatures in the vicinity of 1000 °C for ~200–500 s in an inert Argon environment. The ramp-up and ramp-down rates of the anneal are kept sufficiently high (between 25 and 50 °C/s) so as to induce a thermal stress causing the thin film to disintegrate into a layer of disconnected nanoislands. The control oxide layer is then formed by ALD or CVD techniques. The size, shape, and areal density of the nanocrystals formed in this method can be well controlled by varying the annealing parameters.

In this process, first a thick dielectric layer is grown onto the Si substrate by thermal/CVD techniques to thicknesses of about 30 nm. The dielectric film is then subjected to ion implantation. The typical dose in this method of fabrication is $\sim 1 \times 10^{15}$ to $1 \times 10^{16}/\mathrm{cm}^2$ at energies of few kV (say 1–20 kV), depending upon the required areal density of the ncs and penetration depth. The ion implant process is followed by N_2 ambient annealing at temperatures ~800 °C for about 1 h to enhance the stability of the nc-embedded film. The contacts are formed by E-beam evaporation followed by standard photolithographic processes. In this may leave behind certain amount of defect in the control oxide due to ion bombardment in the implantation process. These defects can, however, give rise to a significant number of bulk charge trapping centers in the control oxide, which could lead to reliability issues. For this reason, a less thicker tunnel oxide, say 15 nm, may be implanted with ncs up to a depth of 5–10 nm and then the control oxide layer may be deposited on top of it making the film more defect free.

Another popular route for the synthesis of nanocrystal-embedded gate dielectric MOS NVMs is the size-selective precipitation technique. The ncs are initially fabricated by solvothermal methods [28]. The synthesized nanocrystals are then dispersed in liquid media (e.g., hexane, methanol) and successively centrifuged to obtain ncs of uniform size distribution. Thereafter by spin coating, the ncs are transferred onto the surface of the tunnel oxide. The volatile suspension media quickly evaporates away and the SiO_2 tunnel oxide is thus coated with ncs of uniform size distribution. To deposit the control oxide on top of the ncs, Plasma-enhanced ALD is employed (Fig. 10).

The most precise control over the size, shape, and the spatial distribution of the ncs can be achieved by self-assembly method. In this technique, a sacrificial template made of porous polymer helps in the growth of nanoislands of specified size, shape, and spacing. The template may be made from phase separation from a dilute solution of a copolymer (like PMMA-PS copolymer) followed by dissolution of one of the polymers in a solvent, as reported by IBM [30]. The porous template is used as a sacrificial layer with a oxide hardmask to etch away SiO_2 and to deposit a-Si in the gaps. In the process, anc-Si–SiO_2 matrix is formed which replicates the

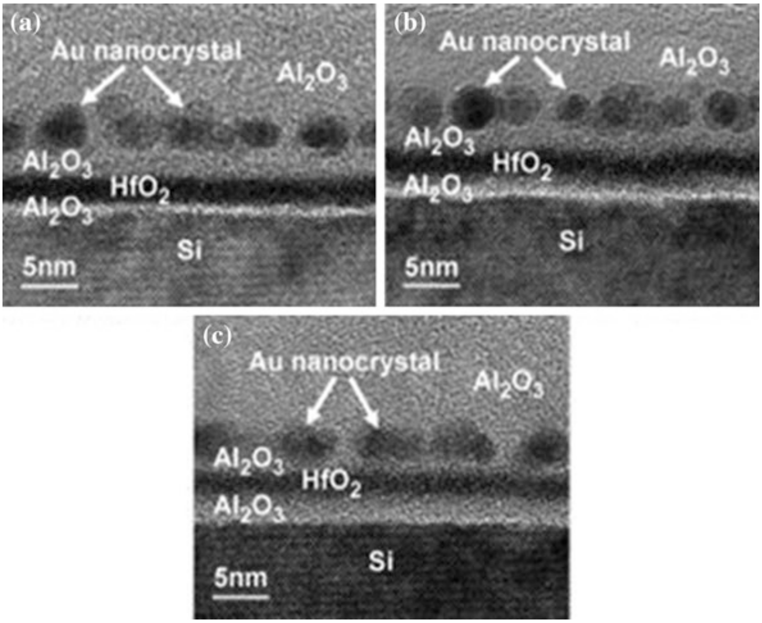


Fig. 10 nc-Au-embedded Al_2O_3–HfO_2 multilayer gate dielectric stacks fabricated by thermal decomposition followed by size-selective precipitation (reprinted with permission from Lo et al. [28]). © AIP 2009

PMMA-PS template in dimensions. Such a method using polymer template-assisted growth of nc-Si has been developed by IBM in 2003 for the fabrication of nc-Si–SiO_2 MOS memory devices [30]. By tailoring the copolymer combinations and template dimensions, one can control the nc-embedded layer dimensions.

2.3 Performance of nc-Embedded MOS NVM Devices

In addition to the prevention of lateral diffusion and better reliability, recent experimental studies have shown the better performance of nc-embedded MOS NVMs, over the conventional FGMOS in terms of various other aspects of NVM like lower write (programming) voltages and lesser direct tunneling leakage currents, higher flatband/threshold voltage shifts, better charge storage properties, and a wider and more stable memory window [9–35] (Fig. 11).

As discussed previously, as the applied gate bias on the MOS NVM is increased gradually, the Fowler–Nordheim and the direct tunneling currents both contribute to the gate current of the device. For lesser gate biases, the contribution of the direct tunneling current is significant, while for larger bias the F–N current tends to dominate. This shifting of gate currents from the predominantly direct tunneling to the F–N tunneling domain can be observed in the gate output characteristic of the MOSFET as a sharp rise of gate current over a certain gate voltage. This particular gate voltage where the F–N tunneling tends to overtake the direct tunneling currents is identified as the minimum write/programming voltage of the MOS NVM. Usually in practice, the device write voltage is kept slightly higher than this value to ensure F–N tunneling of electrons into the charge trap states.

With the introduction of ncs in the gate dielectric, it is seen that this minimum write voltage is sufficiently minimized. In the 30–35 nm thick gate oxide SiO_2–Si MOSCAP devices, fabricated by Ng et al. [13] and Yang et al. [16] it is seen that the minimum write voltage drops by about 25–30 % from its original value of 30 V

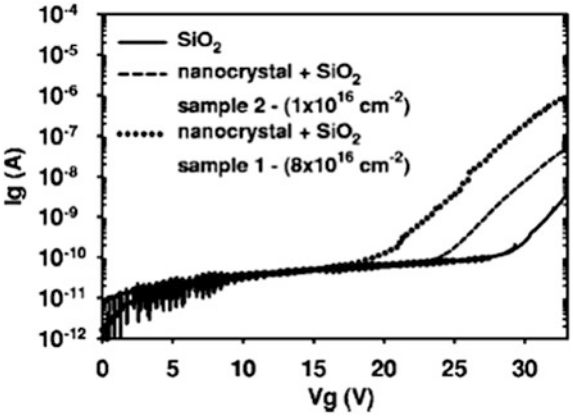


Fig. 11 Lowering of the F–N onset voltage with nc-Si embedding in the gate oxide in MOS NVMs (reprinted with permission from Ng et al. [13]). © AIP 2006

due to nc incorporation in the gate oxide. Such a decrease could be attributed to the lowering of the effective barrier height for electron tunneling in the composite gate dielectric as proposed by Chakraborty et al. [125]. Now it is worth mentioning here that these devices are research purpose devices having considerably thicker dielectrics and Si substrates than industry standard flash. With scaling of the device dimensions and the use of high-k dielectric stacks, the write voltage usually lies in the range 5–7 V [20, 27, 34].

Also since the nanocrystals form a deeper charge trap compared to the conventional FGMOS/SONOS, thus nc-embedded MOS NVMs can store more amount of charge than conventional MOS NVMs. This leads to a higher amount of flatband voltage shift/threshold voltage shift and larger memory windows in such nc-embedded NVMs, which is indicated in the C-V hysteresis curve in experimental studies. Further the stability of the memory window with time is also enhanced by the deeper charge trap of nc-embedded devices as they suppress the direct tunneling leakage from nc to the substrate.

In the early days of nc-embedded MOS NVMs, most reports were on semiconducting nc-embedded MOS memories [9–19]. However, since the early 2000s, the focus has turned toward the various metallic nanocrystals for their higher work function and higher density of states compared to semiconducting ones. Various metallic ncs like Ni [20–22], Ag [21, 23, 25], W [25, 26], Pt [24, 26], Au [22, 24, 27, 28], and Ru [33] have been reported in recent literature. The higher work function of the metallic ncs make for deeper potential wells, and the higher density of states give them the capability of storing more electrons per nanocrystal, than semiconducting ncs [21–28, 31–34]. Another advantage of using metallic ncs is the fact that metal ncs can be synthesized in much smaller sizes (~ 2 nm) than Si or Ge ncs, which enhances the scalability of the NVM device.

In their work, Lee and Meteer [24] have shown the advantages of Au, Ag, and Ptncs compared to the semiconducting Si-ncs-embedded gate oxide MOS NVM devices in terms of the Flatband voltage shift, and a higher and more stable memory window. Such an improvement in the memory window with metallic ncs has also been observed by Lee et al. [34] when comparing nc-Si with nc-Ni-embedded HfO_2 gate dielectric MOS NVM.

Another vital parameter, the leakage current of the MOS memories, has been observed to improve with the application of metallic nc-Ni, instead of nc-Si-embedded gate dielectrics. The barrier heights for electron tunneling across the thin HfO_2 tunnel oxide calculated from experiments by Lee et al. [34] has also increased to 1.93 eV with nc-Ni, from 1.5 eV for nc-Si which explains its better charge retention properties (Table 1).

The improvement in terms of retention time in case of nc-W, and nc-Au was found to be about 100 and 150 times better than nc-Si, respectively, in a study by Li et al. [35].

CNT/C60-embedded gate oxide NVMs: CNT and C60 have also been embedded in the gate dielectric of MOS NVMs to work as charge trap sites. In some reports, these carbon-based nanomaterials have been used in combination with metallic nanocrystals to provide a double layer of charge trap sites in the gate dielectric.

Table 1 Memory effect of various metal ncs

Memory effects of various metal nanocrystals			
Element	Tunnel dielectric	Vth < comma > C–V (V)	Sweep voltage (V)
Au	SiO_2	2.3	2 to −4
Ag	SiO_2	2.1	2 to −4
Pt	SiO_2	3.8	2 to −4
W	SiO_2	0.95	3 to −4
Ni	SiO_2	0.52	6 to −6
	HfO_2	0.75	2 to −2
$NiSi_2$	SiO_2	1.04	3 to −3
	HfO_2	1.38	3 to −3
$CoSi_2$	SiO_2	1.1	3 to −3
Al	AlN	1	5 to −5
$Ni_{1-x}Fe_x$	PI	2	11 to −6
TiN	Al_2O_3	2.9	NA

Used with permission from Yeh et al. [26]. © Elsevier 2007

Fullerenes (C60) could be a very good material for embedding in gate dielectric of MOS NVM devices, due to their small size, chemical stability, and a work function higher than semiconducting ncs. Also C60 does not usually react with the encapsulating oxide which makes it suitable for CMOS applications. Recently, Hou et al. [36, 37] have realized the C60-embedded gate dielectric MOS NVM devices, which have shown better suppression of trap-assisted tunneling of electrons in the presence of C60 in the tunnel oxides, and better flatband shifts and P/E pulse time, compared to those without C60 embedding.

CNT-embedded HfAlO gate oxide MOS memories have also been successfully fabricated by Lu and Dai [76]. These devices have shown better performance in terms of higher capacitance, and lower programming voltages, and faster programming with prominent hole charging effects. However, a major challenge with CNT embedding is the high aspect ratio of CNTs which makes it difficult to prevent them from forming conductive paths within the embedded layer by virtue of touching each other. Such network formation can nullify the advantage of lateral diffusion suppression that is possessed by most nc-embedded MOS NVMs [76].

Multilayer high-k Stacks: One of the major issues in MOS NVM devices, is the direct tunneling leakage through ultrathin tunnel oxides. In this scenario, the use of high-k material like HfO_2, Al_2O_3, HfAlO, and La_2O_3 as gate dielectric in addition to nc embedding is a good way to address the scaling-related leakage. Moreover, as fabrication techniques have advanced, the use of stacked gate oxides comprising of various dielectric material has become quite common for MOS NVMs [21–35]. In most of these stacks, thermal SiO_2 is employed as the bottom layer of the stack for the purpose of better lattice matching with the Si substrate.

A number of recent experimental works have been reported on stacks such as nc-TiW-embedded SiO_2–Al_2O_3 [39] nc-Au-embedded Al_2O_3–HfO_2–Al_2O_3 [28]

stacks, Silicon Rich Silicon dioxide (SRSO)-embedded between HfSiO and SiO_2 or HfO_2 and SiO_2 dielectric layers [40], and TaN/Al_2O_3/HfO_2/SiO_2/Si (TAHOS) [46] stacks.

3 Multigate MOS NVM Devices

From recent studies, the multigate MOSFET architectures, (like Ω-Gate, Π-Gate, Fin FET, and Double-Gate and Gate All Around FETs), have proven to show better performance in terms of higher transconductance, greater suppression of short channel effect (SCE), improved subthreshold swing (SS), and lesser drain-induced barrier lowering (DIBL) compared to the other planar MOS structures. The FinFET has been instrumental in advancing the CMOS VLSI industry to the present 22 nm technology node. Also the better electrostatic control provided by the surround gate, together with the design compatibility with future 3-D IC architectures, make it a very promising structure as well. In this scenario, the exploration of multigate structures for MOS NVM applications is underway. Multigate architectures offer superior electrostatic control of the gate over the Si body which can be used to lower the write voltage and increase the data-retention time (through leakage reduction) MOS NVM devices.

Various multigate memories have been fabricated in recent years, these include the Double-Gate Fin SONOS, body-tied DG SONOS (omega) type NVM, and various surround gate NVMs [41–50]. The channel in these type of NVMs are usually narrow Si Fins or Si/poly-Si nanowires, though sometimes ZnO nanowires have also been reported as channel material [41] (Fig. 12).

In these structures, the gate is wrapped around an etched out Si "Fin" or a nanowire. Thus, the applied gate voltage has the maximum influence on the channel leading to an improved control over the device characteristics. These devices can provide for not only for better leakage resistance and greater immunity to SCEs, but they also provide a larger gate area for storing more charge for lesser amount of device footprint on the chip. Usually, the gate dielectric stacks in such multigate MOS NVM are the oxide–nitride–oxide (ONO) stacks; however, nonconventional advanced stacks like the TAHOS [46] have also been used in some cases. The introduction of nanocrystals within the gate dielectric stack of such a multigate structure is rather challenging; however, there have been some reports on nc-embedded gate dielectric surround gate poly-Si nanowire MOS NVM devices in recent years.

3.1 *Write/Erase Mechanism of Multigate MOS NVM*

The write erase mechanism of such devices is similar to that of conventional SONOS/floating gate MOS NVMs or nc-embedded MOS NVMs. Upon application

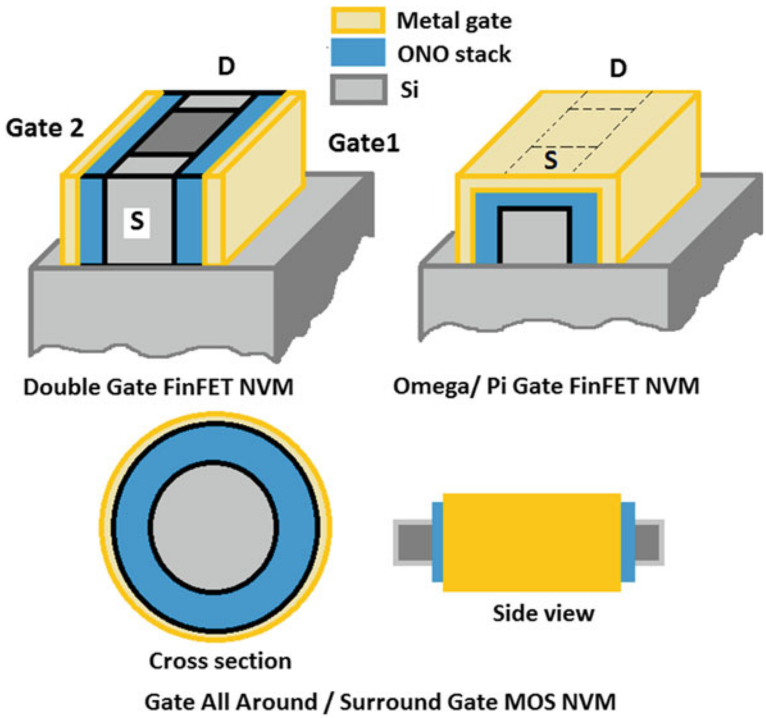
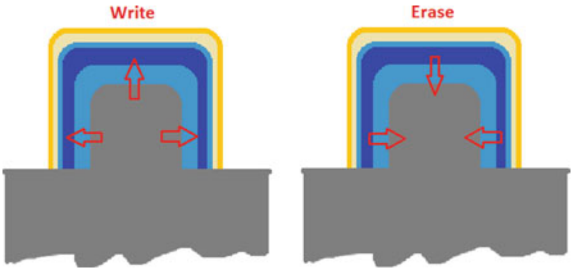


Fig. 12 Different types of architectures in multigate NVMs

of a sufficient gate bias the electrons in the inversion layer in the tunnel oxide-Si interface tunnel from the Si fin/nanowire into the nitride or the embedded nc states by the F–N tunneling mechanism. For erasing a sufficient reverse gate bias causes the electrons trapped in the nitride layer/the nc states to tunnel back into the Si substrate. The fine difference lies in the electrostatics of the multigate channel. From theoretical aspects, there have been a few modeling studies carried out of late on such multigate MOS NVM by Cho et al. [43], Amorosso et al. [126] and Sengupta et al. [127]. These models are based on evaluating charge trapping and detrapping in the nitride trap/the nc states within the composite gate dielectric by applying the WKB approximation of quantum tunneling of electrons. For this purpose, the surface potential of the MOSFET may be assumed fixed above a certain voltage as implemented in the models of Cho et al. [43] and Amorosso et al. [126] or it may be self-consistently evaluated with the charging of the floating gate/nc-embedded layer as performed by Sengupta et al. [127] (Fig. 13).

In the method proposed by Sengupta et al. [127], the electric fields in each of the individual layers of the composite gate dielectric stacks are evaluated for the program/erase modes and from the knowledge of the fields the tunneling currents are

Fig. 13 Write/erase mechanism of a multigate MOS NVM cell



calculated. The process has to be self-consistent with the surface potential evaluation of the MOSFET, since the charge stored in the floating gate/nitride/nc-embedded layer affects the MOS surface potential, which in turn influences the substrate-tunnel oxide barrier height (Fig. 14).

Channel Hot Carrier Injection: This is another write mechanism present in multigate SONOS flash NVM, in this mechanism, electrons moving from source to drain acquire sufficient energy from the applied electric filed (between source and drain) in order to tunnel across the ultrathin tunnel oxide. This usually occurs at source-drain applied field strengths in excess of 100 kV/m, which is quite common for the submicron MOSFETs. A fraction of the energetic electrons are "lucky" enough not to suffer inelastic collisions as they are scattered toward the tunnel oxide can overcome the tunnelling barrier and get into the nc energy states. The probability for such an event is expressed by the so-called "Lucky Electron Model" for channel hot carrier injection and is given by the compound probabilities of the electron gaining sufficient energy [3, 5, 127, 128], that of the electron not suffering collision before reaching the tunnel oxide interface and also that of not suffering any collision in the field reversal region in the oxide (Fig. 15).

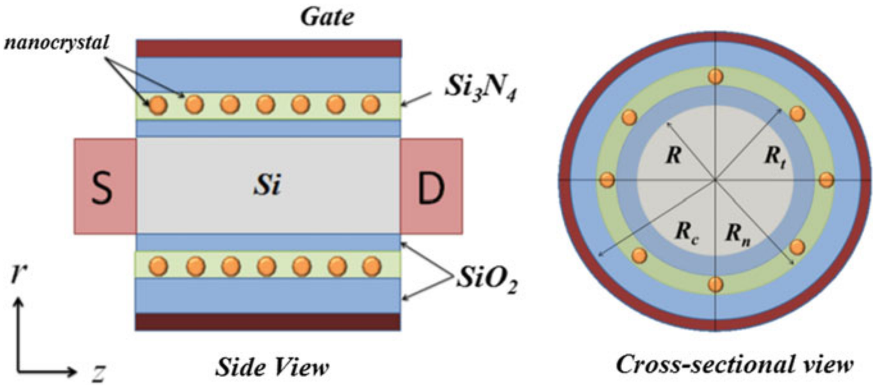


Fig. 14 Schematic diagram (not to scale) of the GAA NVM device proposed

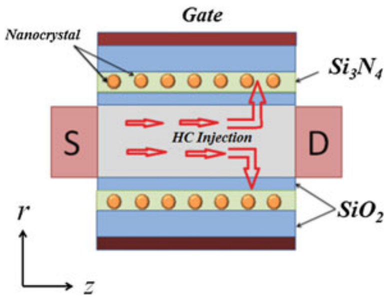


Fig. 15 Schematic of the channel hot carrier injection write process

3.2 Fabrication of Multigate MOS NVM

FinFET/pi Gate NVM: In 2002 Cho et al. [43] reported the omega gate SONOS NVM device of 120 nm channel length. Cho et al. fabricated the Si fins by chemical patterning of Si substrate, followed up by source/drain doping, and sequential deposition of a 2 nm SiO_2 tunnel oxide and a 4 nm thick SiN layer and 1 nm thick SiO_2 control oxide and finally metallization to create the gate electrodes. This process formed the ONO gate dielectric wrapped around in a Ω-shape around the Si fin.

In a very recent work, Yun et al. [50] have fabricated independent double-gate Fin SONOS flash of 90 nm channel length. The fins were patterned by a mix and match photolithography and e-beam lithography process. The ONO stack had thickness of 3.5/12.5/8 nm of tunnel oxide, nitride floating gate, and control oxide, respectively. The deposition technique commonly used for forming the dielectric stacks in such structures is the high-density plasma (HDP) chemical vapor deposition (CVD) process.

GAA nanowire SONOS NVM: The first step for formation of a surround gate nanowire NVM is the fabrication of the nanowire. This can be achieved by pattering of a thin Si substrate into a number of fins using phase-shift mask lithography and dry etch processes. Alternatively poly-Si nanowires can be fabricated by depositing amorphous α-Si by Low Pressure (LP) CVD, followed by solid-phase crystallization in nitrogen ambient and nanowire patterning by electron beam lithography and reactive ion etching (RIE). For ZnO nanowires, there are hydrothermal techniques like chemical bath deposition and also physical methods like vapor–liquid–solid (VLS) deposition. It is important to have a CMOS compatible nanowire fabrication process for future large-scale production of such devices (Fig. 16).

The gate oxide stacks can be then grown sequentially by the LPCVD technique. For the nc-embedded ONO stacks, the ncs are formed by the RTA method. For the nc-embedding process, the nitride layer deposition is performed in steps. First a thin layer of nitride is formed by LPCVD, followed by a thin layer of Si (for nc-Si formation). Thereafter by RTA, the Si film is broken into nc-Si and the final layer of nitride is then deposited on top of it by LPCVD. In this way an nc-Si-embedded nitride layer is formed in the ONO stack.

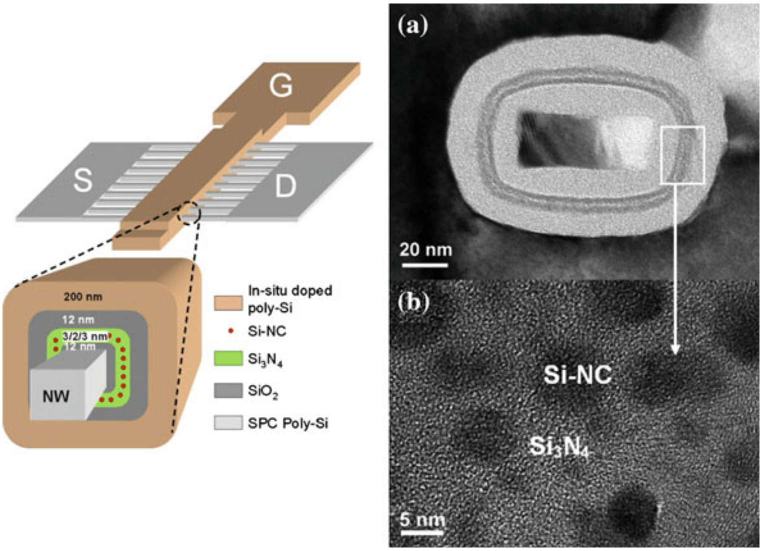


Fig. 16 nc-embedded GAA poly-Si nanowire NVM (with permission from Hung et al. [48]). © AIP 2011

3.3 *Performance of Multigate MOS NVM*

Si nanowire-independent double-gate (iDG) MOSFET NVMs fabricated by Lin et al. [51] have shown good control over the SCEs, and good program/erase efficiency. Also the device has shown suppression of read interrupt signals. However, the device is affected by paired cell interference which means that injected charges in one of the fins influences the opposite by interference coupling.

In 2009, Fu et al. [46] reported a TaN/Al_2O_3/HfO_2/SiO_2/Si (TAHOS) stack-based GAA Si nanowire-based NVM. The nanowire FETs had gate lengths of 300 nm, with DIBL < 50 mV/V and SS < 100 mV/dec. Under 13 V gate stress for 100 µs, the devices show a 3.7 V memory window. The retention characteristics for the said device showed a decrease of 0.2 V in memory window for a waiting time of 10^4 s. The TAHOS stacks used in these cells have shown superior P/E performance and memory windows compared to SONOS stacks in similar GAA Si NW NVMs.

In a very recent work (2011), Hung et al. [48] have reported the memory performance of a SPC poly-Si nanowire GAA MOSFET NVM device, with 2 nm Si-ncs embedded in the nitride layer of 8 nm thickness, in an oxide (12 nm)–nitride (8 nm)–oxide (12 nm) stack. The in situ-doped poly-Si nanowires were rectangular in cross section with a width of 60 nm and a height of 25 nm [48].

The device fabricated by Hung et al. [48] has shown threshold voltage shifts of 2–3 V for write voltage of ±15 V and stress time of 0.1 s, for both electron and hole charging. For the same write voltage and an erase voltage of −23 V for 0.1 s, they tested the device for its retention characteristics. Their studies (by extrapolation method) show that for a 10 year waiting time, the nc-embedded GAA NVM retains 86 % of its initial memory window. This is about 7 % better compared to SONOS GAA NVM devices of similar structures and dimensions.

In another very recent work, Chen et al. [49] have shown a suspended poly-Si nanowire Double-GAA TFT memory device, employing a nc-Si-embedded SiO_2 as the gate dielectric. In their work, they have shown the advantages of having two gate electrodes for their GAA Si nanowire NVM, than one electrode only. The device shows about 83 % memory window retention over an extrapolated period of 10 year. In terms of reliability, the dual GAA device showed 76 % memory window retention after 10^4 program/Erase (P/E) cycles, but the single GAA device broke down after 10^2 P/E cycles.

4 CNT and Graphene-Based NVM

Due to the excellent electrical, thermal, optical, and mechanical properties of CNT and graphene they are being considered as a suitable material to replace the Si CMOS technology in the long run [52]. The problem with graphene is it is zero band-gap that makes large-area graphene FET turn OFF which is unsuitable for logic applications. In this context graphene nanoribbon (GNR) which introduces a small band-gap due to lateral confinement has been proposed as a solution. Also semiconducting CNT with its intrinsic nonzero band-gap could prove to be suitable channel material in the 3-D IC technology. Such developments have prompted researchers to develop NVM devices based on CNT or graphene FET. These devices usually operate on the working principles of the conventional devices like SONOS, or advanced memories like the nc-embedded gate dielectric memories, resistive switching memories, and ferroelectric memories [103–110].

Chan et al. [103] reported back-gated CNT networked channel NVM with Ge nanocrystals embedded in HfO_2 gate dielectric. They fabricated the NVMs on a Si substrate which acts as the back gate electrode. 30 nm HfO_2 was deposited on the Si substrate by atomic layer deposition. On top of this 5 nm diameter Ge nanocrystals embedded in 8 nm HfO_2 formed by PLD technique followed by Cr–Au electrode patterning. The SWCNTs formed by CVD technique separately. Most of the tubes obtained by the process used by Chan et al. were found to be (7, 5) SWNTs in nature. These CNTs were then drop-cast onto the patterned electrodes. The SWCNTs form a conducting network and act as the channel of the back-gated FET.

The device fabricated by Chan et al. [103] works on the same principle as that of the nc-embedded gate dielectric MOS NVM devices. The electrons from the CNT network channel get trapped into the nc states to cause a threshold voltage shift for the NVM.

The device fabricated by them showed a memory window of ~5.3 V for a gate voltage sweep of −5 to +5 V, with a readout conductance ratio (ON/OFF) greater than 10^3. The device showed stability of threshold voltage up to 200 P/E cycles.

Since graphene and CNT are highly flexible material having good optical transparency, they have been explored for the flexible memory applications as well. Such flexible memories could be very useful in wearable electronic gadgets, SMART military uniforms, aerospace applications, flexible displays, and so on. Yu et al. in 2011 [108] have reported highly flexible NVM devices with SWCNT network channel and oxygen decorated graphene electrodes, fabricated on PET substrates. 50 nm ALD grown Al_2O_3 serves as the gate dielectric in such a device. The oxygen atoms attached to the graphene electrodes act as the charge storage nodes in such a memory device. The channel, electrodes, and the substrate all being highly flexible and transparent, the resulting NVM device has shown excellent flexibility and less than 5 % transmission reduction.

The device showed high electron mobility of 44 $cm^2\ V^{-1}\ s^{-1}$ and fast write speeds of 100 ns. Under a gate voltage sweep of ±10 V, the NVM showed a 10 V memory window. The retention time was measured as 1000 s and the endurance characteristics showed stability of the NVM after more than 500 P/E cycles. The mechanical properties were also quite good for more than 1000 cycles of bending tests.

Hwang et al. [106] recently reported a resistive switching NVM having Al/CNT-polystyrine (PS) composite/Al structure with ON/OFF ratios $>10^2$ and P/E endurance $>10^5$, and data retention of more than 10^5. A thin layer of AlO_x acts as the blocking oxide between the CNT-PS and the bottom Al electrode in order to prevent charge leakage from the CNT-PS interface traps. They further investigated tuning of the CNT work function by doping it with B or N and observed that for Al/B-CNT-PS/N-CNT-PS/Al structures have an intermediate stable memory states between the LRS and the HRS of the standard RRAM. The NVM devices fabricated on polyamide substrates were highly flexible as well.

Such multilevel memories are capable of storing more than 2-bits per cell as in standard two level RRAMs.

Among graphene/CNT-based FerroFET NVM, the one developed by Fu et al. [105] with PECVD grown 1 nm dia SWCNT channel with $BaTiO_3$ ferroelectric and Nb-doped $SrTiO_3$ back gate showed ultra-low power NVM operation. The energy consumption in the write process was 0.1 fJ for a 600 nm channel length device. The readout ON and OFF currents were 40 and 0.02 nA for write voltages of −6 and +6 V and the drain voltage of 200 mV. Song et al. also reported a robust bistable single-layer graphene (SLG) FeFET having a 180 nm PZT layer as the ferroelectric material and Pt/Ti back gate. Zheng et al. [109] have successfully fabricated a SLG FeFET with P(VDF-TrFe) ferroelectric layer. The device showed a very high electron mobility of 700 $cm^2\ V^{-1}\ s^{-1}$. The device also showed very good $\Delta R/R$ ratio of 3.5.

5 Nanomechanical/Mass Transport Memories

Another very interesting class of NVM devices is being explored recently which uses electromechanical actuation to store information. The prominent among them are the NEMS (Nanoelectromechanical System) memories and the reversible mass transport or CNT nanoparticle shuttle memories [111–116].

The NEMS NVMs mostly work on the principle of charging and discharging a floating gate in a MOS device by means of mechanical contact and separation (Fig. 17).

Usually, the charge is stored in a metallic floating gate or nanocrystal-embedded layer of the gate dielectric stack of a MOSFET. Unlike the conventional floating gate NVMs where the electrons tunnel from the Si substrate into the floating gate/the nanocrystal energy states, in the NEMS NVM the charging and discharging is done by a metallic cantilever/suspended beam which can contact with the floating gate upon actuation. The cantilever/suspended beam is generally made of multiple metallic layers for mechanical durability and fast actuation. A voltage is applied to the metal cantilever/beam and the opposite polarity voltage is applied to the actuating electrode to bring it into contact with the floating gate of the MOSFET. Depending upon the bias on the cantilever/beam, the floating gate can get positively or negatively charged. The charge on the floating gate controls the conductivity of the MOSFET channel, and therefore by sensing the drain current of the MOSFET two distinguishable readout states "0" and "1" can be achieved.

In another kind of NEMS NVM architecture, a suspended nanocrystal-embedded SiO_2 beam suspended in a cavity serves as the charge storage node.

With the application of a pulse negative write voltage in the control gate, the ncs get negatively charged. Now as the beam is actuated upwards the decrease in negative gate bias on the MOSFET channel gives a higher drain current, which can be considered as the ON state of the device. However, if the beam is actuated

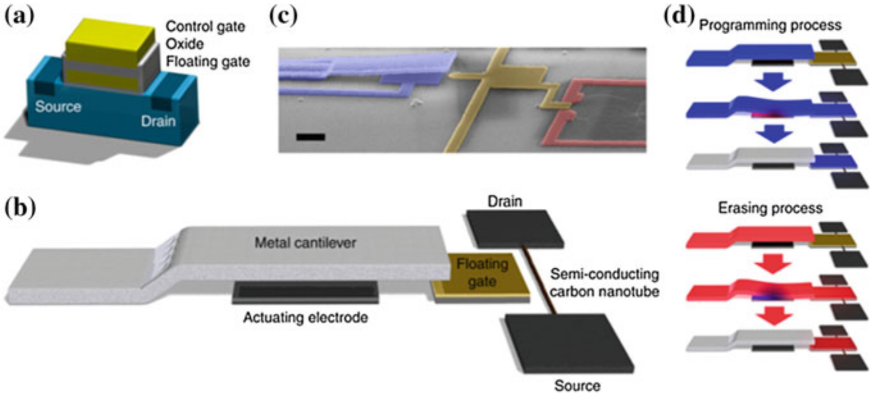


Fig. 17 Schematic diagram and the operation of a cantilever-type floating gate NEMS NVM (reprinted with permission from Lee et al. [115]). © Nature Publishing Group 2011

downwards, the negative gate bias on the MOSFET channel increases and a low current OFF state is achieved. Clearly, the read operation is based on the sensing of the MOSFET drain current. For device operations, the sidewalls of the cavity are grounded.

The fabrication processes involved in forming these NEMS NVM are a combination of standard CMOS fabrication process for forming the MOSFET and specialized surface/bulk micromachining and etching techniques like Deep-reactive ion etching (DRIE) and LIGA for creating structures such as nanocantilevers, clamped switches and cavities.

Lee et al. [115] fabricated a Cr/Au/Cr triple-layer cantilever which controlled the charge on an Au floating gate of a semiconducting CNT FET with Al_2O_3 gate dielectric. Their results showed a readout ON/OFF ratio greater than 10^5, with 11 h repeated switching operations without any significant change in the ON state and the OFF state currents. The endurance was measured to be ~ 500 P/E cycles. The currents in the ON and the OFF states were measured to be 10^{-8}–10^{-12} A [115].

Jang et al. [113] fabricated a cantilever type and a clamped switch type NEMS NVM device using conventional CMOS fabrication technology, having a beam thickness of 30 nm and an air gap between the suspended bridge/cantilever and the floating gate being 20 nm. They observed a ON/OFF ratio of $\sim 10^5$ a subthreshold slope (of bit switching) to be about 3 mV/dec and showed endurance of several hundred P/E cycles.

Also Jang et al. [112] have developed a NEMS capacitor NVM based on vertically aligned CNTs. These devices have shown, very low leakages, high switching speeds of up to 750 MHz, and low power consumption.

The chief advantage of NEMS NVM lies in their very good leakage resistance and data-retention properties. However, structures are much complex and slower than standard phase change (PC) NVMs like the RRAM, FeRAM, STT-MRAM, etc., and therefore there application as mainstream NVM devices may not be very easy.

Another type of NVM structure based on nanomanipulation technique, has been demonstrated recently. A CNT reversible mass transport memory, also known as a CNT shuttle memory is such a device where a metallic nanoparticle/nanorod of Fe or Co is encapsulated in a CNT and its position within the CNT is manipulated by means of electrostatic actuation [111] (Fig. 18).

Depending upon the applied electrode bias, the position of the Fe nanorod or the nanoparticle in the CNT can be controlled accurately. Depending upon the position of the core (i.e., Fe nanorod/nanoparticle) the device can be in a number of states (three-level memory in case of the nanorod core and multilevel for nanoparticle core). The readout operation for the NVM fabricated by Bergtrup et al. [111] is accomplished by measuring the resistance of the CNT which varies with the position of the Fe nanoparticle. This is because of the fact that the Fe nanoparticle despite remaining fully encapsulated within the CNT all time for this type of device, acts as a position-dependent scattering center, and thereby creates distinguishable position-dependent readout resistances.

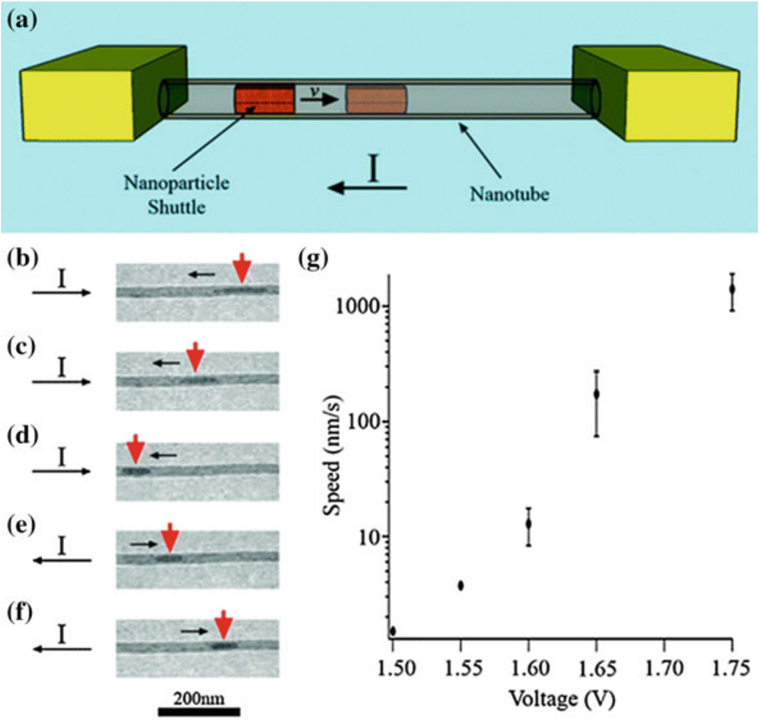


Fig. 18 CNT-nanocrystal shuttle memory SEM (reprinted with permission from Begtrup et al. [111]). © ACS 2009

The most important step in the fabrication of such devices lies in the formation of the core-shell nanostructure consisting of Fe nanoparticle/nanorod and CNT. Bergtrup et al. [111] carried out this by a single-step process involving the pyrolysis of Ferrocene in argon environment at 1000 °C. This gave core-shell Fe-CNT structures, which were then transferred to the substrate in isopropanol media. The Pd electrodes were thereafter patterned using standard photolithographic deposition and lift-off techniques.

Such devices have shown to operate at lower voltages than most conventional and advanced NVMs. Also the switching speeds of the CNT mass transport mechanism has been experimentally reported to be in the range of $\sim 10^{-11}$ s [116] which is sufficiently fast. It may be further lowered by the optimization of the CNT diameter, length, and the dimensions of the shuttle (Fe). Also theoretically such NVMs can have a possible memory density of up to 1 Tb/in^2 and since such CNT core-shell structures are thermodynamically highly stable, hence such an NVM is theorized to be stable over a billion years [111].

Such devices, though only at research stage, at present, could play an important role in the long run in the field of archival data storage due to their multilevel nature, high projected capacity, and material stability.

6 Conclusion

With conventional floating gate and SONOS MOS NVM devices facing serious challenges owing to scaling in the VLSI sector, several advanced memory technologies have entered the race to be the future replacement of Si MOS NVMs. Among them the prominent are the nanocrystal-embedded memories, resistive switching memories, ferroelectric FETs, STT-RAM, and so on. Also advanced multigate architectures like the double gate, FinFET, and surround gate/nanowire FETs have also come to play to try and extend the reign of MOS NVM devices. Novel 3-D structures like the NEMS memory or the CNT nanoparticle shuttle memories are also being investigated by researchers.

All these emerging memories have some pros and cons, while resistive switching/PC NVMs have the best scalability, lowest cost, and very good 3-D integration for making high-density chips they are left wanting in the area of endurance. In this area, the STT NVM or the Ferro NVM score over the PC NVM devices. However, in terms of scalability and 3-D stacking FeRAM is left behind. The current feature size of the present charge trap MOS NVM flash devices is 22 nm with ITRS projected feature size in 2024 being 8 nm. In this context, the PCM devices are presently having a 45 nm feature size with scalability up to 8 nm. FeRAM and STT-RAM, however, have this limit only up to 65 and 16 nm, respectively. However, what makes these devices extremely useful is their speed. While the best MOS flash NVMs of today has read, write, and erase times of 0.1, 1, and 0.1 ms, respectively, without any solutions of further improvement by 2024, the PCM, STT, and FeNVMs have shown read, write, and erase speeds below 50 ns and are projected to become even faster (down to 10 ns) by 2024, as the technology evolves. In terms of retention times, all these emerging NVM devices exceed the current MOS flash NVMs. As per the endurance issue the FeRAM, STT-RAM, and PCM devices show endurance of 10^{15}, 10^{12}, and 10^{9} P/E cycles at present. These values indicate that although the FeRAM, STT-RAM, and PCM has a lesser endurance than current DRAM (10^{16} P/E cycles), these devices can be used as a replacement as the NAND flash NVM for data storage application as the present flash NVMs are capable handling only 10^{4} P/E cycles, with ITRS projections for 5×10^{3} cycles in 2024.

As for the novel structures like the NEMS NVM or the CNT mass transport memory, such devices have a projected scalability only down to 50 nm, which is not sufficient considering current NAND flash to have feature sizes of 22 nm. However, they could offer lesser write energies compared to flash NVMs and could find applicability in specific sectors like sensor systems and long-term archival memories.

In their recent work, Smullen et al. [129] have compared the overall performance index of flash, STT, and the PCRAM devices from the point of view of speed/performance, energy requirements, leakage, memory density, endurance, and retention time. It is evident from such studies that the PCM devices clearly cover all the requirements of the present NVM memory technology with performance superior to NAND flash NVM in all aspects. Therefore, these emerging memory devices such as Ferro NVM, PCM, STT memories have already established themselves as suitable candidates in terms of scalability, speed, data retention, and endurance to replace the charge trap MOS NVM cells, currently employed in most NAND flash architectures, and ongoing research on the fabrication, structural improvements, and intensive studies into the underlying physics of these devices would provide a new generation of NVM devices with superior speed, higher memory density, improved data retention, and reliability at lower costs.

References

1. S.E. Thompson et al., A 90-nm logic technology featuring strained-silicon. IEEE Trans. Electron Dev. **51**(11), 1790–1797 (2004)
2. J.T. Kedzierski, A Decade of nanoelectronics journey from classical bulk CMOS to metal-gate FinFETs, in *IEEE EDS DLT* (Kolkata and personal communication) 17 Nov 2011
3. P. Cappelletti, C. Golla, P. Olivo, E. Zanoni (eds.), *Flash Memories* (Kluwer Academic, Boston, MA, 1999)
4. I.S. Stievano (ed.), *Flash Memories* (InTech, Rijeka, Croatia, 2011)
5. R. Micheloni, L. Crippa, A. Marelli (eds.), *Inside NAND Flash Memories* (Springer, New York, 2010)
6. A.K. Sharma, *Semiconductor Memories: Technology, Testing and Reliability* (Wiley, Hoboken, NJ, 1997)
7. G. Campardo, R. Micheloni, D. Novosel, *VLSI-design of Non-Volatile Memories* (Springer, Berlin, Heidelberg, 2005)
8. R. Bez, E. Camerlenghi, A. Modelli, A. Visconti, Introduction to flash memory. Proc. IEEE **91**(4), 489–502 (2003)
9. S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E.F. Crabbé, K. Chan, A silicon nanocrystals based memory. Appl. Phys. Lett. **68**(10), 1377 (1996)
10. J.-D. Choe, K.H. Yeo, Y.J. Ahn, J.J. Lee, Low voltage program/erase characteristics of Si nanocrystal memory with damascene gate FinFET on bulk Si wafer. J. Semicond. Technol. Sci **6**(2), 68–73 (2006)
11. Y. Liu, T.P. Chen, C.Y. Ng, M.S. Tse, P. Zhao, Y.Q. Fu, S. Zhang, Random capacitance modulation due to charging/discharging in Si nanocrystals embedded in gate dielectric. Nanotechnology **16**, 1119–1122 (2005)
12. C.Y. Ng, T.P. Chen, L. Ding, M. Yang, J.I. Wong, P. Zhao, X.H. Yang, K.Y. Liu, M.S. Tse, A.D. Trigg, S. Fung, Influence of Si nanocrystal distributed in the gate oxide on the MOS capacitance. IEEE Trans. Electron Dev. **53**, 730–736 (2006)
13. C.Y. Ng, T.P. Chen, P. Zhao, L. Ding, Y. Liu, Electrical characteristics of Si nanocrystal distributed in a narrow layer in the gate oxide near the gate synthesized with very-low-energy ion beams. J. Appl. Phys. **99**, 106105 (2006)
14. C.Y. Ng, T.P. Chen, M. Yang, J.B. Yang, L. Ding, C.M. Li, A. Du, A. Trigg, Impact of programming mechanisms on the performance and reliability of nonvolatile memory devices based on Si nanocrystals. IEEE Trans. Electron Dev. **53**, 663–667 (2006)
15. C.Y. Ng, T.P. Chen, A. Du, A study of the influence of tunnel oxide thickness on the performance of flash memory based on ion-beam synthesized silicon nanocrystals. Phys. Stat. Sol. A **203**(6), 1291–1295 (2006)
16. M. Yang, T.P. Chen, Z. Liu, J.I. Wong, W.L. Zhang, S. Zhang, Y. Liu, Effect of annealing on charge transfer in Ge nanocrystal based nonvolatile memory structure. J. Appl. Phys. **106**, 103701 (2009)

17. R. Beyer, J.V. Borany, H. Burghardt, Microelectronic engineering interface and border trap relaxation in Si–SiO_2 structures with Ge nanocrystals examined by transient capacitance spectroscopy. Microelectron. Eng. **86**(7–9), 1859–1862 (2009)
18. M. Yang, T.P. Chen, W. Zhu, J.I. Wong, S. Zhang, Comparison of charge storage behavior of electrons and holes in a continuous Ge Nanocrystal layer. Nanosci. Nanotechnol. Lett. **2**, 7–10 (2010)
19. W.K. Choi, W.K. Chim, Observation of memory effect in germanium nanocrystals embedded in an amorphous silicon oxide matrix of a metal–insulator–semiconductor structure. Appl. Phys. Lett. **80**(11), 2014–2016 (2002)
20. J.J. Lee, D.-L. Kwong, Metal nanocrystal memory with high-k tunneling barrier for improved data retention. IEEE Trans. Electron Dev. **52**(4), 507–511 (2005)
21. S.-W. Ryu, J.-W. Lee, J.-W. Han, S. Kim, Y.-K. Choi, Designed workfunction engineering of double-stacked metal nanocrystals for nonvolatile memory application. IEEE Trans. Electron Dev. **56**(3), 377–382 (2009)
22. W. Guan, S. Long, M. Liu, Z. Li, Y. Hu, Q. Liu, Fabrication and charging characteristics of MOS capacitor structure with metal nanocrystals embedded in gate oxide. J. Phys. D Appl. Phys. **40**(9), 2754–2758 (2007)
23. S.-W. Ryu, C.B. Mo, S.H. Hong, Y.-K. Choi, Nonvolatile memory characteristics of NMOSFET With Ag nanocrystals synthesized via a thermal decomposition process for uniform device distribution. IEEE Trans. Nanotechnol. **7**(2), 145–150 (2008)
24. C. Lee, J. Meteer, V. Narayanan, E.C. Kan, Self-assembly of metal nanocrystals on ultrathin oxide for nonvolatile memory applications. J. Electron. Mater. **34**(1), 1–11 (2005)
25. Y. Pei et al., MOSFET nonvolatile memory with a high-density tungsten nanodot floating gate formed by self-assembled nanodot deposition. Semicond. Sci. Technol. **24**(4), 045022 (2009)
26. P. Yeh, L. Chen, P. Liu, D. Wang, T. Chang, Metal nanocrystals as charge storage nodes for nonvolatile memory devices. Electrochimica Acta **52**(8), 2920–2926 (2007)
27. V. Mikhelashvili et al., Nonvolatile low-voltage memory transistor based on SiO_2 tunneling and HfO_2 blocking layers with charge storage in Au nanocrystals. Appl. Phys. Lett. **98**(21), 212902 (2011)
28. Y.-S. Lo, K.-C. Liu, C.-W. Cheng, J.-Y. Wu, C.-H. Hou, Field enhancement effect of nanocrystals in bandgap engineering of tunnel oxide for nonvolatile memory application. Appl. Phys. Lett. **94**, 082901 (2009)
29. R. Ang, T.P. Chen, M. Yang, J.I. Wong, M.D. Yi, The charge trapping and memory effect in SiO_2 thin films containing Ge nanocrystals. J. Phys. D Appl. Phys. **43**(1), 015102 (2010)
30. http://www.03.ibm.com/press/us/en/pressrelease/22083.wss
31. J.V. Borany et al., The formation of narrow nanocluster bands in Ge-implanted SiO_2-layers. J. Appl. Phys. **43**, 1159–1163 (1999)
32. S. Koliopoulou, P. Dimitrakis, D. Goustouridis, P. Normand, C. Pearson, M.C. Petty, H. Radamson, D. Tsoukalas, Metal nano-floating gate memory devices fabricated at low temperature. Microelectron. Eng. **83**, 1563–1566 (2006)
33. S.-S. Yim, M.-S. Lee, K.-S. Kim, K.-B. Kim, Formation of runanocrystals by plasma enhanced atomic layer deposition for nonvolatile memory applications. Appl. Phys. Lett. **89** (9), 093115 (2006)
34. J.J. Lee, D.-L. Kwong, Metal nanocrystal memory with high-tunneling barrier for improved data retention. IEEE Trans. Electron Dev. **52**(4), 507–511 (2005)
35. Y. Li, S. Liu, Using different work function nanocrystal materials to improve the retention characteristics of nonvolatile memory devices. Microelectron. J. **40**(1), 92–94 (2009)
36. T.-H. Hou, U. Ganguly, E.C. Kan, Programable molecular orbital states of C60 from integrated circuits. Appl. Phys. Lett. **89**(25), 253113 (2006)
37. T-H. Hou et al., in *Device Research Conference, 2008* (Santa Barbara, CA, 23–25 June 2008), pp. 275–276
38. X.B. Lu, J.Y. Dai, Memory effects of carbon nanotubes as charge storage nodes for floating gate memory applications. Appl. Phys. Lett. **88**(11), 113104 (2006)

39. S. Yang, Q. Wang, M. Zhang, S. Long, J. Liu, Titanium–tungsten nanocrystals embedded in a SiO_2/Al_2O_3 gate dielectric stack for low-voltage operation in non-volatile memory. Nanotechnology **21**, 245201 (2010)
40. L. Khomenkova, B.S. Sahu, A. Slaoui, F. Gourbilleau, Hf-based high-k materials for Si nanocrystal floating gate memories. Nanoscale Res. Lett. **6**(1), 172 (2011)
41. D. Yeom et al., ZnO nanowire-based nano-floating gate memory with Ptnanocrystals embedded in Al_2O_3 gate oxides. Nanotechnology **19**(39), 395204 (2008)
42. F. Hofmann et al., NVM based on FinFET device structures. Solid State Electron. **49**, 1799–1804 (2005)
43. I.H. Cho, B.G. Park, J.D. Lee, Nano-scale SONOS memory with a double-gate MOSFET structure. J. Kor. Phys. Soc. **42**(2), 233–236 (2003)
44. S.W. Oh, S.S. Park, J.H. You, T.W. Kim, H.W. Kim, Nanoscale two-bit NAND silicon-oxide-nitride-oxide-silicon flash memories with a pair of double gate FinFET structures. J. Kor. Phys. Soc. **55**(1), 263–266 (2009)
45. K.H. Yuen, T.Y. Man, S. Member, A.C.K. Chan, M. Chan, S. Member, A 2-bit MONOS nonvolatile memory cell based on asymmetric double gate MOSFET structure. IEEE Electron. Dev. Lett. **24**(8), 518–520 (2003)
46. J. Fu et al., Integration of high-dielectrics and metal gate on gate-all-around Si-nanowire-based architecture for high-speed nonvolatile charge-trapping memory. IEEE Electron. Dev. Lett. **30**(6), 662–664 (2009)
47. M. Chen, H.Y. Yu, N. Singh, Y. Sun, N.S. Shen, Vertical-Si-nanowire SONOS memory for ultrahigh-density application. IEEE Electron. Dev. Lett. **30**(8), 879–881 (2009)
48. M.-F. Hung, Y.-C. Wu, Z.-Y. Tang, High-performance gate-all-around polycrystalline silicon nanowire with silicon nanocrystals nonvolatile memory. Appl. Phys. Lett. **98**(16), 162108 (2011)
49. H.-B. Chen, Y.-C. Wu, C.-K. Yang, L.-C. Chen, J.-H. Chiang, C.-Y. Chang, Impacts of poly-Si nanowire shape on gate-all-around flash memory with hybrid trap layer. IEEE Trans. Electron Dev. **32**(10), 1382–1384 (2011)
50. J. Yun, S. Member, Y. Kim, I.H. Park, J.H. Lee, D. Kang, M. Lee, H. Shin, S. Member, J.D. Lee, B. Park, A. Fin, Independent double-gate fin SONOS flash memory fabricated with sidewall spacer patterning. IEEE Trans. Electron Devices **56**(8), 1721–1728 (2009)
51. H. Lin, Z. Lin, W. Chen, T. Huang, Read characteristics of independent double-gate poly-Si nanowire SONOS devices. IEEE Trans. Electron Dev. **58**(11), 3771–3777 (2011)
52. International Technology Roadmap for Semiconductors 2011 Report, Emerging Research Devices (ERD)
53. http://www.hpl.hp.com/news/2010/jul-sep/memristorhynix.html
54. S. Kim, H. Moon, D. Gupta, S. Yoo, Y. Choi, Resistive switching characteristics of sol-gel zinc oxide films for flexible memory applications. IEEE Trans. Electron Dev. **56**(4), 696–699 (2009)
55. D.L. Lewis, H.H.S. Lee, Architectural evaluation of 3D stacked RRAM caches, in *2009 IEEE International Conference on 3D System Integration* (Sep 2009), pp. 1–4
56. D. Lee et al., Resistance switching of the nonstoichiometric zirconium oxide for nonvolatile memory applications. IEEE Electron Dev. Lett. **26**(10), 719–721 (2005)
57. F. Zhuge, S. Peng, C. He, X. Zhu, X. Chen, Y. Liu, R.-W. Li, Improvement of resistive switching in Cu/ZnO/Pt sandwiches by weakening the randomicity of the formation/rupture of Cu filaments. Nanotechnology **22**(27), 275204 (2011)
58. F. Zhuge, B. Hu, C. He, X. Zhou, Z. Liu, R.-W. Li, Mechanism of nonvolatile resistive switching in graphene oxide thin films. Carbon **49**(12), 3796–3802 (2011)
59. W.-Y. Chang, C.-A. Lin, J.-H. He, T.-B. Wu, Resistive switching behaviors of ZnO nanorod layers. Appl. Phys. Lett. **96**(24), 242109 (2010)
60. B. Hu, F. Zhuge, X. Zhu, S. Peng, X. Chen, L. Pan, Q. Yan, R.-W. Li, Nonvolatile bistable resistive switching in a new polyimide bearing 9-phenyl-9H-carbazole pendant. J. Mater. Chem. **22**(2), 520 (2012)

61. P. Misra, A.K. Das, L.M. Kukreja, Switching characteristics of ZnO based transparent resistive random access memory devices grown by pulsed laser deposition. Physica Status Solidi (C) **7**(6), 1718–1720 (2010)
62. M. Lee, Y. Park, B. Kang, S. Ahn, C. Lee, K. Kim, G. Stefanovich, J. Lee, S. Chung, Y. Kim, C. Lee, J. Bong, I. Baek, I. Yoo, 2-Stack 1D-1R cross point structure with oxide diodes as switch elements for high density resistance RAM applications. IEDM technical digest, in *IEEE International Electron Devices Meeting* (2007), pp. 771–774
63. M. Rozenberg, I. Inoue, M. Sánchez, Nonvolatile memory with multilevel switching: a basic model. Phys. Rev. Lett. **92**(17), 178302 (2004)
64. D.V.M.G. Dearnaley, A.M. Stoneham, Electrical phenomena in amorphous oxide films. Rep. Prog. Phys. **1129**(33), 1129–1191 (1970)
65. H. Sim, D. Choi, D. Lee, S. Seo, M. Lee, I. Yoo, H. Hwang, Resistance-switching characteristics of memory application. IEEE Electron Dev. Lett. **26**(5), 292–294 (2005)
66. Q. Tian, M. Tang, F. Jiang, Y. Liu, J. Wu, R. Zou, Y. Sun, Z. Chen, R. Li, J. Hu, Large-scaled star-shaped α-MnS nanocrystals with novel magnetic properties. Chem. Commun. **47**(28), 8100–8102 (2011)
67. Y.H. Tseng, W.C. Shen, C.J. Lin, Modeling of electron conduction in contact resistive random access memory devices as random telegraph noise. J. Appl. Phys. **111**(7), 73701–737015 (2012)
68. J. Won Seo, S.J. Baik, S.J. Kang, Y.H. Hong, J.H. Yang, K.S. Lim, A ZnO cross-bar array resistive random access memory stacked with heterostructure diodes for eliminating the sneak current effect. Appl. Phys. Lett. **98**(23), 233505 (2011)
69. Lee, Hesse, APL **80**(6), 1040 (2002)
70. W. Fu, Z. Xu, X. Bai, C. Gu, E. Wang, Intrinsic memory function of carbon nanotube-based ferroelectric field-effect transistor. Nano Lett. **9**(3), 921–925 (2009). http://pubs.acs.org/doi/abs/10.1021/nl801656w
71. J. Hoffman, X. Pan, J.W. Reiner, F.J. Walker, J.P. Han, C.H. Ahn, T.P. Ma, Ferroelectric field effect transistors for memory applications. Adv. Mater. **22**(26–27), 2957–2961 (2010)
72. S.-Y. Chen, C.-L. Sun, S.-B. Chen, A. Chin, Bi 3.25 La 0.75 Ti_3O_{12} thin films on ultrathin Al_2O_3 buffered Si for ferroelectric memory application. Appl. Phys. Lett. **80**(17), 3168 (2002)
73. S. Das, J. Appenzeller, FeTRAM. An organic ferroelectric material based novel random access memory cell. Nano Lett. **11**(9), 4003–4007 (2011)
74. H. Ishiwara, Memory operations of 1T2C-type ferroelectric memory cell with excellent data retention characteristics. IEEE Trans. Electron Dev. **48**(9), 2002–2008 (2001)
75. M.A. Khan, U.S. Bhansali, H.N. Alshareef, High-performance non-volatile organic ferroelectric memory on banknotes. Adv. Mater. **24**, 2165–2170 (2012)
76. Y.T. Lee, P.J. Jeon, K.H. Lee, R. Ha, H.-J. Choi, S. Im, Ferroelectric nonvolatile nanowire memory circuit using a single ZnO nanowire and copolymer top layer. Adv. Mater. **24**(22), 3020–3025 (2012)
77. F. Mfis, U. Plzt, S.T.O. Si, Nonvolatile memory operations of metal-ferroelectric insulator-semiconductor (MFIS) FET's using PLZT/STO/SI(100) structures. IEEE Electron Dev. Lett. **18**(4), 160–162 (1997)
78. S.R. Rajwade, S. Member, K. Auluck, J.B. Phelps, K.G. Lyon, J.T. Shaw, E.C. Kan, S. Member, A ferroelectric and charge hybrid nonvolatile memory—part i: device concept and modeling. IEEE Trans. Electron Dev. **59**(2), 441–449 (2012)
79. S.R. Rajwade, S. Member, K. Auluck, J.B. Phelps, K.G. Lyon, J.T. Shaw, E.C. Kan, S. Member, A ferroelectric and charge hybrid nonvolatile memory—part II: experimental validation and analysis. IEEE Trans. Dev. Mater. Reliab. **59**(2), 450–458 (2012)
80. G.A. Salvatore, Ferroelectric field effect transistor for memory and switch applications. *Small* **4990** (2011)
81. E.B. Song, B. Lian, S. Min Kim, S. Lee, T.-K. Chung, M. Wang, C. Zeng, G. Xu, K. Wong, Y. Zhou, H.I. Rasool, D.H. Seo, H.-J. Chung, J. Heo, S. Seo, K.L. Wang, Robust bi-stable memory operation in single-layer graphene ferroelectric memory. Appl. Phys. Lett. **99**(4), 042109 (2011)

82. Y. Zheng, G.-X. Ni, C.-T. Toh, M.-G. Zeng, S.-T. Chen, K. Yao, B. Özyilmaz, Gate-controlled nonvolatile graphene-ferroelectric memory. Appl. Phys. Lett. **94**(16), 163505 (2009)
83. Z. Diao, D. Apalkov, M. Pakala, Y. Ding, A. Panchula, Y. Huai, Spin transfer switching and spin polarization in magnetic tunnel junctions with MgO and AlOx barriers. Appl. Phys. Lett. **87**(23), 232502 (2005)
84. Z. Diao, Z. Li, S. Wang, Y. Ding, A. Panchula, E. Chen, L.-C. Wang, Y. Huai, Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory. J. Phys. Condens. Matter **19**(16), 165209 (2007)
85. Z. Diao, M. Pakala, A. Panchula, Y. Ding, D. Apalkov, L.-C. Wang, E. Chen, Y. Huai, Spin-transfer switching in MgO-based magnetic tunnel junctions (invited). J. Appl. Phys. **99** (8), 08G510 (2006)
86. Z. Diao, A. Panchula, Y. Ding, M. Pakala, S. Wang, Z. Li, D. Apalkov, H. Nagai, A. Driskill-Smith, L.-C. Wang, E. Chen, Y. Huai, Spin transfer switching in dual MgO magnetic tunnel junctions. Appl. Phys. Lett. **90**(13), 132508 (2007)
87. G.D. Fuchs, N.C. Emley, I.N. Krivorotov, P.M. Braganca, E.M. Ryan, S.I. Kiselev, J.C. Sankey, D.C. Ralph, R.A. Buhrman, J.A. Katine, Spin-transfer effects in nanoscale magnetic tunnel junctions. Appl. Phys. Lett. **85**(7), 1205 (2004)
88. Y. Huai, Spin-Transfer Torque MRAM (STT-MRAM): challenges and prospects. AAPPS Bull. **18**(6), 33–40 (2008)
89. Y. Huai, F. Albert, P. Nguyen, M. Pakala, T. Valet, Observation of spin-transfer switching in deep submicron-sized and low-resistance magnetic tunnel junctions. Appl. Phys. Lett. **84** (16), 3118 (2004)
90. Y. Huai, M. Pakala, Z. Diao, D. Apalkov, Y. Ding, A. Panchula, Spin-transfer switching in MgO magnetic tunnel junction nanostructures. J. Magn. Magn. Mater. **304**(1), 88–92 (2006)
91. Y. Huai, M. Pakala, Z. Diao, Y. Ding, Spin transfer switching current reduction in magnetic tunnel junction based dual spin filter structures. Appl. Phys. Lett. **87**(22), 222510 (2005)
92. T. Kawahara, Challenges toward gigabit-scale spin-transfer torque random access memory and beyond for normally off, green information technology infrastructure (invited). J. Appl. Phys. **109**(7), 07D325 (2011)
93. P. Krzysteczko, X. Kou, K. Rott, A. Thomas, G. Reiss, Current induced resistance change of magnetic tunnel junctions with ultra-thin MgO tunnel barriers. J. Magn. Magn. Mater. **321** (3), 144–147 (2009)
94. C.J. Lin, S.H. Kang, Y.J. Wang, K. Lee, X. Zhu, W.C. Chen, X. Li, W.N. Hsu, Y.C. Kao, M. T. Liu, M. Nowak, N. Yu, 45 nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell, in *2009 IEEE International Electron Devices Meeting* (*IEDM*) (Dec 2009), pp. 1–4
95. M. Nakayama, T. Kai, N. Shimomura, M. Amano, E. Kitagawa, T. Nagase, M. Yoshikawa, T. Kishi, S. Ikegawa, H. Yoda, Spin transfer switching in TbCoFe/CoFeB/MgO/CoFeB/TbCoFe magnetic tunnel junctions with perpendicular magnetic anisotropy. J. Appl. Phys. **103**(7), 07A710 (2008)
96. S.S.P. Parkin, C. Kaiser, A. Panchula, P.M. Rice, B. Hughes, M. Samant, S. Yang, Giant tunnellingmagnetoresistance at room temperature with MgO (100) tunnel barriers. Nat. Mater. **3**(December), 2–7 (2004)
97. Y. Qi, D. Xing, J. Dong, Relation between Julliere and Slonczewski models of tunneling magnetoresistance. Phys. Rev. B **58**(5), 2783–2787 (1998)
98. C.W. Smullen, V. Mohan, A. Nigam, S. Gurumurthi, M.R. Stan, Relaxing non-volatility for fast and energy-efficient STT-RAM caches, in *2011 IEEE 17th International Symposium on High Performance Computer Architecture* (Feb 2011), pp. 50–61
99. C.W. Smullen, V. Mohan, A. Nigam, S. Gurumurthi, M.R. Stan, Relaxing non-volatility for fast and energy-efficient STT-RAM caches, in *2011 IEEE 17th International Symposium on High Performance Computer Architecture* (Feb 2011), pp. 50–61
100. S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, K. Ando, Fe/MgO/Fe magnetic tunnel junctions. Nat. Mater. **3**(December), 868–871 (2004)

101. H. Zhao, A. Lyle, Y. Zhang, P.K. Amiri, G. Rowlands, Low writing energy and sub nanosecond spin torque transfer switching of in-plane magnetic tunnel junction for spin torque transfer random access memory. J. Appl. Phys. **720**(2011), 2011–2014 (2013)
102. W. Zhao, E. Belhaire, C. Chappert, F. Jacquet, P. Mazoyer, New non-volatile logic based on spin-MTJ. Physica Status Solidi (a) **205**(6), 1373–1377 (2008)
103. M.Y. Chan, L. Wei, Y. Chen, L. Chan, P.S. Lee, Charge-induced conductance modulation of carbon nanotube field effect transistor memory devices. Carbon **47**(13), 3063–3070 (2009)
104. G. Fei, G. Li, L. Wu, H. Xia, A spatially and temporally controlled shape memory process for electrically conductive polymer–carbon nanotube composites. Soft Matter **8**(19), 5123 (2012)
105. W. Fu, Z. Xu, X. Bai, C. Gu, E. Wang, Intrinsic memory function of carbon nanotube-based ferroelectric field-effect transistor. Nano Lett. **9**(3), 921–925 (2009)
106. S.K. Hwang, J.M. Lee, S. Kim, J.S. Park, H.I. Park, C.W. Ahn, K.J. Lee, T. Lee, S.O. Kim, Flexible multilevel resistive memory with controlled charge trap B- and N-doped carbon nanotubes. Nano Lett. **12**(5), 2217–2221 (2012)
107. E.B. Song, B. Lian, S. Min Kim, S. Lee, T.-K. Chung, M. Wang, C. Zeng, G. Xu, K. Wong, Y. Zhou, H.I. Rasool, D.H. Seo, H.-J. Chung, J. Heo, S. Seo, K.L. Wang, Robust bi-stable memory operation in single-layer graphene ferroelectric memory. Appl. Phys. Lett. **99**(4), 042109 (2011)
108. W.J. Yu, S.H. Chae, S.Y. Lee, D.L. Duong, Y.H. Lee, Ultra-transparent, flexible single-walled carbon nanotube non-volatile memory device with an oxygen-decorated grapheneelectrode. Adv. Mat. Deerfield Beach Fla. **23**(16), 1889–1893 (2011)
109. Y. Zheng, G.-X. Ni, C.-T. Toh, M.-G. Zeng, S.-T. Chen, K. Yao, B. Özyilmaz, Gate-controlled nonvolatile graphene-ferroelectric memory. Appl. Phys. Lett. **94**(16), 163505 (2009)
110. F. Zhuge, B. Hu, C. He, X. Zhou, Z. Liu, R.-W. Li, Mechanism of nonvolatile resistive switching in graphene oxide thin films. Carbon **49**(12), 3796–3802 (2011)
111. G.E. Begtrup, W. Gannett, T.D. Yuzvinsky, V.H. Crespi, A. Zettl, Nanoscale reversible mass transport for archival memory 2009. Nano Lett. **9**(5), 1835–1838 (2009)
112. J.E. Jang, S.N. Cha, Y.J. Choi, D.J. Kang, T.P. Butler, D.G. Hasko, J.E. Jung, J.M. Kim, G. A.J. Amaratunga, Nanoscale memory cell based on a nanoelectromechanical switched capacitor. Nature Nanotechnol. **3**(1), 26–30 (2008)
113. W.W. Jang, J.-B. Yoon, M.-S. Kim, J.-M. Lee, S.-M. Kim, E.-J. Yoon, K.H. Cho, S.-Y. Lee, I.-H. Choi, D.-W. Kim, D. Park, NEMS switch with 30 nm-thick beam and 20 nm-thick air-gap for high density non-volatile memory applications. Solid State Electron. **52**(10), 1578–1583 (2008)
114. J.W. Kang, Q. Jiang, Electrostatically telescoping nanotube nonvolatile memory device. Nanotechnology **18**(9), 095705 (2007)
115. S.W. Lee, S.J. Park, E.E.B. Campbell, Y.W. Park, A fast and low-power microelectromechanical system-based non-volatile memory device. Nature Commun. **2**, 220 (2011)
116. J. Zhao, J.-Q. Huang, F. Wei, J. Zhu, Mass transportation mechanism in electric-biased carbon nanotubes. Nano Lett. **10**(11), 4309–4315 (2010)
117. Q.-A. Huang, J.K.O. Sin, M.C. Poon, Field emission from silicon including continuum energy and surface quantization. Appl. Surf. Sci. **119**, 229–236 (1997)
118. R.H. Fowler, L. Nordheim, Electron emission in intense electric fields, in *Proceedings of the Royal Society of London*, vol. 119 (1928), pp. 173–181
119. M. Lenzlinger, E.H. Snow, Fowler Nordheim tunneling into thermally grown SiO_2. J. Appl. Phys. **40**, 278–283 (1969)
120. E.L. Murphy, R.H. Good Jr, Phys. Rev. **102**, 1464 (1956)
121. J.G. Simmons, Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. J. Appl. Phys. **34**(6), 1793–1803 (1963)
122. J. Maserjian, N. Zamani, Behavior of the Si/SiO_2 interface observed by Fowler-Nordheim tunneling. J. Appl. Phys. **53**(1), 559–567 (1981)

123. K.-S. Wen, H.-H. Li, C.-Y. Wu, A new gate current simulation technique considering Si/SiO$_2$ interface trap generation. Solid State Electron. **38**(4), 851–859 (1995)
124. N. Matsuo, Y. Takami, Y. Kitagawa, Modeling of direct tunneling for thin SiO$_2$ film on n-type Si (1 0 0) by WKB method considering the quantum effect in the accumulation layer. Solid State Electron. **46**(4), 577–579 (2002)
125. G. Chakraborty, A. Sengupta, F.G. Requejo, C.K. Sarkar, Study of the relative performance of silicon and germanium nanoparticles embedded gate oxide in metal-oxide-semiconductor memory devices. J. Appl. Phys. **109**, 064504 (2011)
126. S.M. Amoroso, C.M. Compagnoni, A. Mauri, A. Maconi, A.S. Spinelli, A.L. Lacaita, Semi-analytical model for the transient operation of gate-all-around charge-trap memories. IEEE Trans. Electron. Dev. **58**(9), 3116–3123 (2011)
127. A. Sengupta, *Studies on Nanoelectronic Devices.* Ph.D. thesis, Jadavpur University (2012)
128. S. Tam, P.-K. Ko, C. Hu, Lucky-electron model of channel hot-electron injection in MOSFET'S. IEEE Trans. Electron Dev. **31**(9), 1116–1125 (1984)
129. C.W. Smullen et al., Relaxing non-volatility for fast and energy-efficient STT-RAM caches, in *2011 IEEE 17th International Symposium on High Performance Computer Architecture* (Feb 2011), pp. 50–61

Nanocrystalline Thin Film Gas Sensors

Amretashis Sengupta

1 Introduction

The term sensor comes from the Latin word '*sentire*' which means to perceive [1]. In electronics sensors are a type of devices that converts some non-electrical input parameter (which we want to measure) into electrical signals having some correlation with the magnitude and nature of the input. Sensors have invaded every sphere of modern industry. In industrial automation, consumer electronics, automobile, space-exploration, medical sector, sensors are everywhere. One of the chief applications of sensors is for the detection of gases and chemical vapors. These class of sensors known as chemical/gas sensors are of prime importance when it comes to industrial control, automation in chemical industry, quality control in food processing and agricultural industry, personnel safety in potentially hazardous environments (like mines, outer space), military applications (in case of chemical warfare and homeland security), biomedical sciences (for disease detection and treatment) and many more. Sensing of an explosive gas like methane is one such application which can be helpful in ensuring safety in coal mines and in industries where LNG/CNG fuel (which is constituted of predominantly methane) is used.

Of the various types of Gas sensors available in the market MEMS (Micro Electro Mechanical System) based (also called "chip-type" sensors by some) have the advantage of low cost, low power, batch production, and the possibility of use with ICs [2–5].

It has been a well-established trend to use metal oxides for gas sensing owing to their low cost, small size, and ease of operation. Different oxides like SnO_2, WO_3, ZnO, and In_2O_3 have been in use for long time and they have been thoroughly

---

A. Sengupta (✉)
School of VLSI Technology, Indian Institute of Engineering Science and Technology,
Shibpur, Howrah 711103, India
e-mail: a.sengupta@vlsi.iiests.ac.in

© Springer-Verlag Berlin Heidelberg 2015

A. Sengupta and C.K. Sarkar (eds.), *Introduction to Nano*,
Engineering Materials, DOI 10.1007/978-3-662-47314-6_9

studied for their sensitivity toward different gases [2–7]. Of these the focus of this work has been on methane sensing by Zinc Oxide.

Though many a material (oxides of tin, tungsten, titanium, etc.) have been reported for the application in the field of sensing explosive gases there have been very few reported applications in ZnO for such applications [8–14]. The use of ZnO thin films for sensing a reducing gas hydrogen using DC resistive technique was first reported in 2005 by [8]. The fabrication of a thin film ZnO methane sensor fabricated using sol–gel technique was first reported in 2006 by Bhattacharyya et al. [9].

In recent years, there has been a keen interest shown by researchers to investigate the gas sensing properties of nanostructured materials (like rods, fibers, tubes) because of their high surface to volume ratio and expected reduction in activation energy for chemisorption. Zinc oxide is one such material which has become a well-established sensor material owing to its easy synthesis procedure, good electrical properties and in particular its compatibility with Si leading to MEMS-based chemical sensors [14]. Many reports regarding the sensing of various gases and gas mixtures like H_2 [15, 16], CO [17], NO_2 [18], ozone [15], acetone vapor [19], ethanol vapor [20], LPG [21, 22] using ZnO nanostructures have been published recently. Due to its high surface to volume ratio the gas sensing properties of aligned ZnO nanorods are markedly better compared to their conventional thin film counterparts and they have the advantages of greater sensitivity, faster response, and lower operating temperatures (which is essentially useful in lowering the power requirements for possible integration with CMOS technology). There exist many synthesis techniques for ZnO nanorods. Of them hydrothermal [23], solvothermal [24], solution phase deposition [25], pulsed laser deposition [26], CVD, and CBD (Chemical Bath Deposition) [27, 28] are widely adapted. Of these methods CBD provides a low temperature, fast, versatile, and easily controllable method for the synthesis of various ZnO nanostructures on a variety of substrates. The advantage of this method also includes the possibility of large-scale production of such structures and the low energy requirement. Though there have been a number of reports regarding methane sensing with sol–gel grown or electro deposited ZnO thin films [8–14], however methane sensing with CBD grown ZnO nanorods still remains somewhat less explored.

2 General Overview of Sensors

Human beings like all living organisms interact with their environment in various ways. We see, hear, touch, smell, or taste various objects and gather information about them and process this information in our brain to draw inference and in this way we “sense” things. A detailed study of the mechanisms involved in our senses reveals that there exist various types of receptors in our sense-organs (namely the rod cells and cone cells in our retina that respond to light intensity and frequency, Corty’s organ in our ear that respond to mechanical vibration, Meissner cells that give the sense of pressure, i.e., touch). In this sense all these receptors of different stimuli can be termed as sensors. But in the perspective of the electronic world the word "Sensor" has somewhat different dimension to it.

3 Basic Definitions

The Instrument society of America defines a sensor as "*a device which provides a usable output in response to a specified measurand*" [1, 29, 30].

There is however a thin line of difference between a sensor and a transducer; whereas a sensor converts some non-electrical parameter and responds with an electrical signal which has some form of quantitative correlation to the non-electrical parameter the action of a transducer is merely that of conversion of one form of signal to another. So we may say all sensors are transducers but not all transducers are sensors [3, 30].

The variations in the output electrical signal parameter are a function of the input parameter. The physical quantity such as temperature, pressure, some chemical concentration, etc. measured by the sensor (/transducer) is called the *measurand*. The portion of the sensor (/transducer) in which the electrical output signal originates is the *sensing element (/transduction element)*. The upper and the lower values of the measurand for which the sensor produces useful output are known as the *dynamic range* of the sensor [29].

4 Sensing Principles

Several basic transduction elements can be used in sensors for different measurands. They include capacitive, electromagnetic, electrochemical, inductive, resistive, photoconductive, photovoltaic, and piezoelectric elements. Sensors are largely *nonlinear sensors* but with appropriate electronics they are designed to provide an output which is a linear function of the measurand.

Different effects, different principles, and different materials are applied for making the device, keeping in mind what we want to measure and how accurately

Table 1 Sensors classified on measurand

Domain	Electrical property/effect
Mechanical	Piezoelectricity, Piezoresistivity, capacitance
Thermal	Seebek effect, Peltier effect, pyroelectric effect, thermoelectric effect
Magnetic field	Hall effect, magnetoresistance
Radiant energy	Photovoltaic, photoelectric
Chemical (incl. biological)	Conductometry, calorimetry

Table 2 Sensors classified on sensing principle

Principle/effect	Sensors
Piezoelectricity and piezoresistivity	Force, shear, vibration, SAW devices, thermistors
Pyroelectricity	IR sensors, thermal imaging
Magnetism	GMR sensors, spintronic devices
Conductivity/potentiometry	Gas sensor, Ion sensing FETs, humidity sensors, bio-sensors, breath analyzer
Optical effect	SPR (Including plasmonics), Fiber optic sensors

we want to measure. Avery brief flavor of the variety of sensors based on their working principle is given in Tables 1 and 2.

5 Sensor Classification

There exist myriad ways to classify Sensors, they may be based on Transduction principle, measurand, materials or even the technology, Applications or simply based on the properties of the sensor. Possibly the most common way of classification is the one which classifies sensors into active and passive sensors. *Active sensors* require an external source signal source/excitation whereas *Passive sensors* are capable *of* generating their own electrical output signal. Thus strain gages, thermistors can be classified as active sensors while thermocouples, photodiodes are passive sensors.

6 Sensor Characteristics

Parameters that are useful in order to evaluate just how good a sensor is are known as the sensor characteristics. Some of them are listed below (*the most important ones being the '4S's—Sensitivity, Selectivity, Speed, and Stability*) [1, 5].

Transfer Function: For any sensor, the quantity to be sensed (measurand) and the sensor output has some sort of functional relationship. If the measurand is given by x then the output would be related to it as $y = f(x)$, this function $f(x)$ is known as the *Transfer Function* [1, 3, 30].

Full Scale Output (FSO): The upper limit of the output for a sensor is called the Full scale (FS) [1, 3, 30].

Sensitivity: Mathematically it is the slope of the calibration curve, which means a measure of how the output varies for a small change in the input (measurand). The Sensitivity at a point x_a is given by $S(x_a) = \frac{dy}{dx}\big|_{x=x_a}$ [1, 3, 30].

Speed: This is measured by how quickly the sensor can respond to changes in measurand and how fast it can get back to 90 % of its original output value once the measurand is removed [1, 3, 30].

Selectivity: It is the ability of a sensor to measure the target measurand in presence of several others and environmental factors. For example, a CH_4 sensor will be said to be selective if its output for CH_4 is much more in comparison to other gases present in a mixture and is very little affected by Relative Humidity [1, 3, 30].

Stability: The ability of a sensor is to behave in the same way in response to measurand, and to maintain the same input–output characteristics. If there is a drift in the characteristics over time the sensor is deemed less-stable [1, 3, 30].

Resolution: The minimal change of the measurand value which can be detected successfully by the sensor is called the resolution [1, 3, 30].

Accuracy: It is a measure of how correct the output of the sensor under consideration is. Basically, it is comparison with the response of a standard sensor under the same operating conditions [1, 3, 30].

Linearity/NonLinearity: Ideally, we want our sensor to have a linear transfer function, e.g., $f(x) = kx + b$ but this is hardly the case in most real-life applications and the response is nonlinear in nature (e.g., some polynomial of x).

Offset: Output of the sensor with zero input (measurand) applied. An example is the dark current in a photo-diode.

A few more points that are worth mentioning from the industry point of view are [2, 4].

Manufacturability: This is very important, since a sensor is not just for research purposes but for the real-life application. If the manufacturing process is too long term, or too complex it may not be industrially viable.

Cost: If the cost of the sensor is too much then it will not cut in the market and fail to compete with other manufacturers. So minimizing the cost by understanding the Techno-economics is another aspect.

Aesthetics/Packaging: It is another important part of sensor manufacturing. Depending upon the target market the packaging should be such that the sensor offers best possible performance in that environment yet it is convenient to put it into the desired system and easy to maintain or replace.

7 Gas Sensors

Gas sensors are devices that respond to the presence of gas molecules (by some form of transduction action) and gives out electrical signal commensurate with the gas concentration [1, 5].

8 Fundamentals of Gas Sensing

Before going into the details of gas sensors, some of the commonly used terms encountered during discussion of gas sensors need to be explained as given below:

Parts per million (ppm): It can be calculated from the following relation:

$$C = \frac{V_g}{V_a + V_g}$$

where C is the concentration in volume fraction V_g is the volume of the gas, V_a the volume of air. To obtain concentration as a percentage, C is to be multiplied by 100 and to get ppm and ppb (parts per billion); C is to be multiplied by 10^{-6} and 10^{-9}, respectively.

For flammable gases that mix with air and combust upon exposure certain trigger/ignition source (spark, flame, etc.) the ***Lower Explosive Limit (LEL)*** is the minimum required concentration of gas in air so it may be ignited. The ***Upper Explosive Limit (UEL)*** is the maximum concentration of gas in the mixture to permit ignition. If concentration is above UEL there is insufficient oxygen presence to aid successful ignition. For example, the LEL and UEL of methane are 5 and 15 %, respectively. For toxic gases there are certain exposure limits for personnel working in such environments. The **Short-Term Exposure Limit *(STEL)*** is the maximum possible concentration permissible for a period of 10 min (time weighed average). For **Long-Term Exposure Limit *(LTEL)*** is the maximum possible concentration permissible for a period of 8 h (time weighed average). For carbon monoxide, the STEL and LTEL are 300 and 50 ppm, respectively [2, 4].

In the field of chemical sensing, the focus of the gas monitoring systems revolves around the following:

(i) Oxygen monitoring for health and combustion engines, boilers, etc.
(ii) Inflammable gases like methane, LPG, CNG, hydrogen) monitoring to prevent hazards.
(iii) For toxic gases (i.e., gases like carbon monoxide, sulfur dioxide, chlorine) and automobile exhaust pollutants like SO_x, KNO_x, etc.).
(iv) Relative humidity is an important parameter in weather monitoring and climate control. Also it can impact the readings of the aforementioned target gases.

Owing to this myriad application range the demand of low cost, miniaturized, low power, reliable, and widely deployable gas sensors has increased in the market. This has triggered huge research and development in the field of solid state sensors, especially metal-oxide gas sensors.

9 Various Types of Gas Sensors

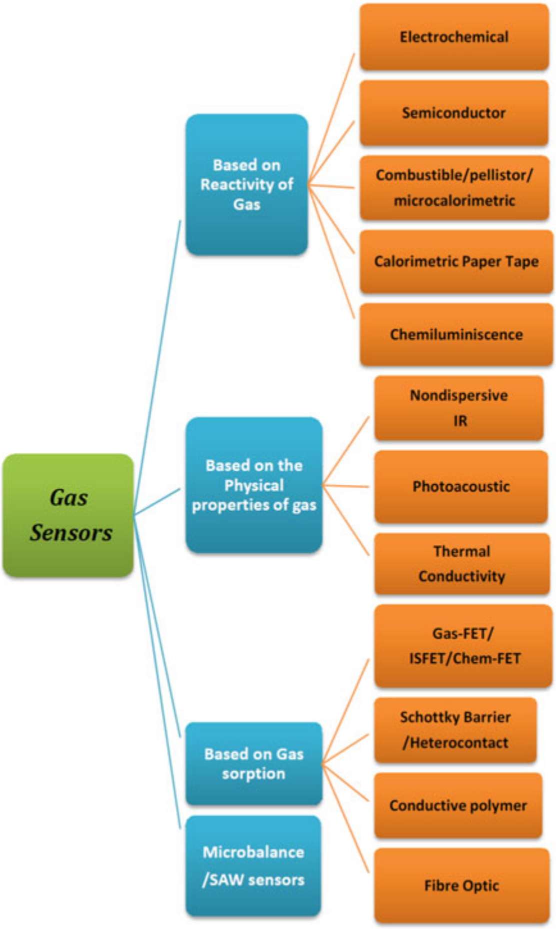
There exist various types of gas sensors based on the sensing principle (Ref: Fig. 1). Of these the Semiconductor gas sensor is of our particular interest due to their low cost, low power consumption, miniaturization prospect, and robustness [2, 4, 5]. The challenges for these type of sensors lie in selectivity and characteristics drift, its performance can be greatly improved by techniques such as sensitization and application of nanostructured materials [6, 9–12].

Design-wise semiconductor most sensors are of the three types mentioned below

Taguchi Type

Named after their inventor N. Taguchi, Japan (invented in 1968); this type of sensors (short name TGS) was the first commercially produced Solid state sensor and are very widely used till date. A coating of gas sensitive composition is made

Fig. 1 Various type of gas sensor based on working principle [1, 5]



on the outer surface of an alumina tube (having few millimeters of outer diameter and length and less than a millimeter thick). The electrodes and lead wires are taken out from the ends. A small resistance heater coil is placed inside the tube so as to get an operating temperature of 300–400 °C on the sensing layer. The power requirement of TGS is around 1 W. Though SnO_2 is the most commonly used sensing material in this type of sensors. Also recently the application of new materials such as nanomaterials, CNTs are being used to improve TGS [2, 4, 6].

Printed Sensors

The power requirement can be reduced significantly (down to few hundred mW) by using printed sensor. In this type the sensor coating is given on one surface of a thin alumina substrate and a resistive heater is screen printed on the reverse side [2, 4].

Chip Type/MEMS-Based Sensors

MEMS-based sensor are chiefly used for low power (few mW) application and IC integration (where the sensor and associated circuit is fabricated on a single chip). Here micro-heaters are patterned by lithography on thin membranes of Si that leads to better thermal efficiency. As the power consumption in a metal-oxide gas sensor is chiefly due to the heating arrangements required, the use of MEMS heaters can greatly improve the power consumption. Also the miniaturization wise these sensors are highly lucrative [2, 4, 5] (Fig. 2).

In brief we may summarize the requirements of a “good” gas sensor as a fairly high degree of sensitivity, stability, speed, selectivity (the essential 4S) coupled with reliability, ease of manufacturing and the necessary techno-economics of cost and aesthetics. The types and applications of sensors are vast and ever changing and an exhaustive discussion on that matter is beyond the scope of this work. So, the focus here will be limited to ZnO-based methane sensor.

10 Semiconductor Gas Sensors

When it comes to semiconductor gas sensors we mostly speak of oxide materials. Different oxides like SnO_2, WO_3, ZnO, and In_2O_3 have been in use for long time and they have been thoroughly studied for their sensitivity toward different gases [7]. The technique used here is mostly DC conductometric (i.e., measuring the change in DC current or the change in resistance of an oxide thin film which is modulated accordingly with the varying gas concentrations in the ambient).

11 Basic Mechanism of Semiconductor Oxide Gas Sensors

The basic schematic is shown in Fig. 3 where the target gas is allowed to fall on a thin film of a porous oxide material. The film is usually deposited on an oxidized Si substrate. The fabrication aspect and other technical details will be discussed later.

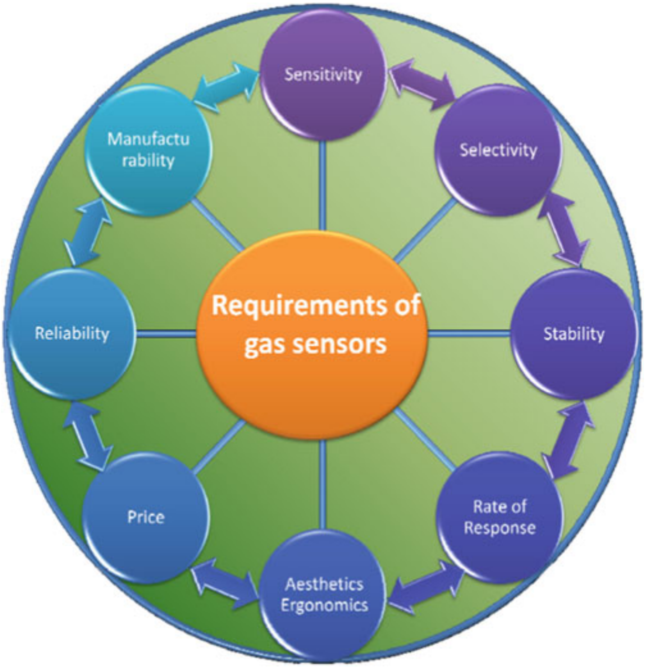


Fig. 2 Basic requirements of a gas sensor [3–5]

The most important phenomena in thin film sensors occur at the exposed oxide surface. Due to poor coordination between anions and cations in such oxides there exist number of dangling bonds on the surface, which are effective in adsorbing gaseous species. So there exist certain gas sensitive regions in the granular surface microstructure (Fig. 4) [6].

From the ambient air Oxygen, adsorbs on the oxide surface. As shown in Fig. 5.

It is notable that the oxygen which is adsorbed on the surface of the oxide film dissociates and then pulls the electrons from the surface toward itself due to its strong electron affinity. Thus, it eventually forms a layer of O^- on the surface. As the electrons are pulled out from the conduction band a depletion layer of thickness

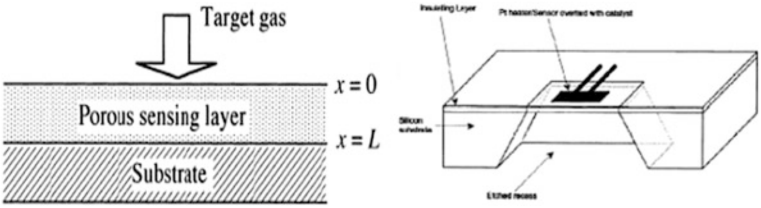


Fig. 3 A basic thin film gas sensor schematic (reprinted with permission from Ref. [6]) © Springer LLC 2003

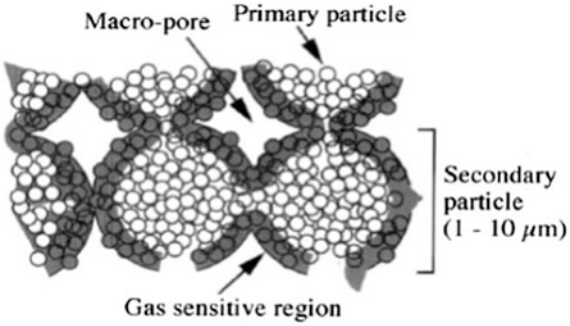


Fig. 4 The gas sensitive regions on porous film surface (*top view*) (reprinted with permission from Ref. [6]) © Springer LLC 2003

L (also called the Debye length) is thus formed due to depletion of free charge carriers (conduction band electrons in case of n-type metal oxides such as ZnO). So now apart from the grain boundary barrier, an electron also encounters the depletion layer as it moves on the surface of the thin film. As a result there is a decrease in conductivity of the film. The grain size has a significant impact on the sensitivity of thin films as we shall see in the next section.

Now as a reducing gas like Hydrogen comes into contact with the surface it breaks upinto H^+ and then reacts with the adsorbed O^- and produces H_2O. In the process it donates two electrons back to the oxide surface, thus restoring the conductivity. The result of this effect can be easily understood from the sharp increase in conductivity as the gas is exposed. When the gas is removed the surface is again depleted of carriers as the oxygen is again adsorbed on the surface and the carriers depleted [5–7] (Fig. 6).

The entire reaction of breaking up of a reducing gas into H^+ takes place at elevated temperatures (>100 °C) and thus the water vaporizes away immediately.

The resistance goes down for increasing concentration of the reducing gas and the response *S* is the highest for a particular temperature called the operating temperature. The response time, operating temperature, the values of the resistance of the film, etc. all depend on the sensing material and the target gas combination.

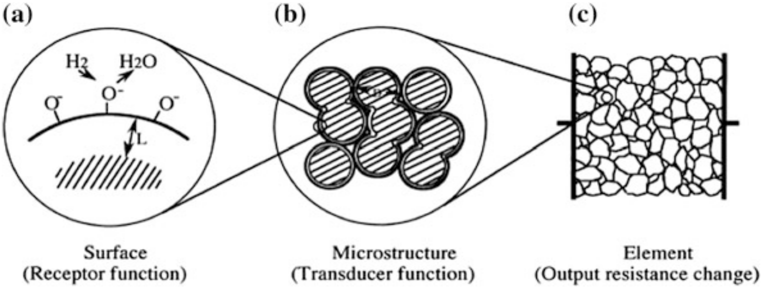


Fig. 5 Oxygen adsorption on sensor surface (reprinted with permission from Ref. [6]) © Springer LLC 2003

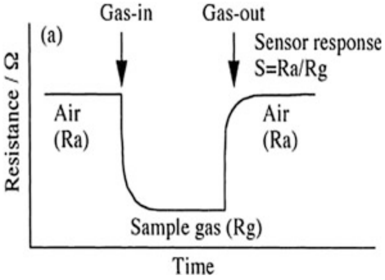


Fig. 6 Ideal sensor response to a reducing (reprinted with permission from Ref. [6]) © Springer LLC 2003

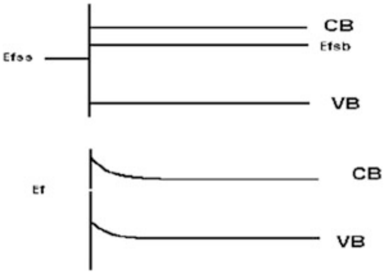


Fig. 7 Explanation by band diagram (as Fermi levels of surface and bulk equalize band bending takes place which is subsequently removed by the action of a reducing gas)

This can also be explained by the band structure as follows; As oxygen as O^- are adsorbed in the surface, they take away electrons thus lowering the Fermi level of surface EFSS (in comparison to bulk EFB) and as a result potential barriers are formed due to band bending in addition to the already existent inter-grain barriers. Then the reducing gas dissociates and produces H^+ which reacts with adsorbed Oxygen at elevated temperature to produce H_2O which desorbs immediately also giving electrons in the process which removes the band bending effects and restores conductivity [1] (Fig. 7).

12 Effect of Particle Size and the Nano Advantage

Now as mentioned earlier the particle size plays a very important role in gas sensitivity. The grain size effect on the sensitivity of metal-oxide gas sensors has been explained by Xu et al. [31] with a semi-quantitative model. This is schematically shown in Fig. 8a–c. According to their model necks are formed between

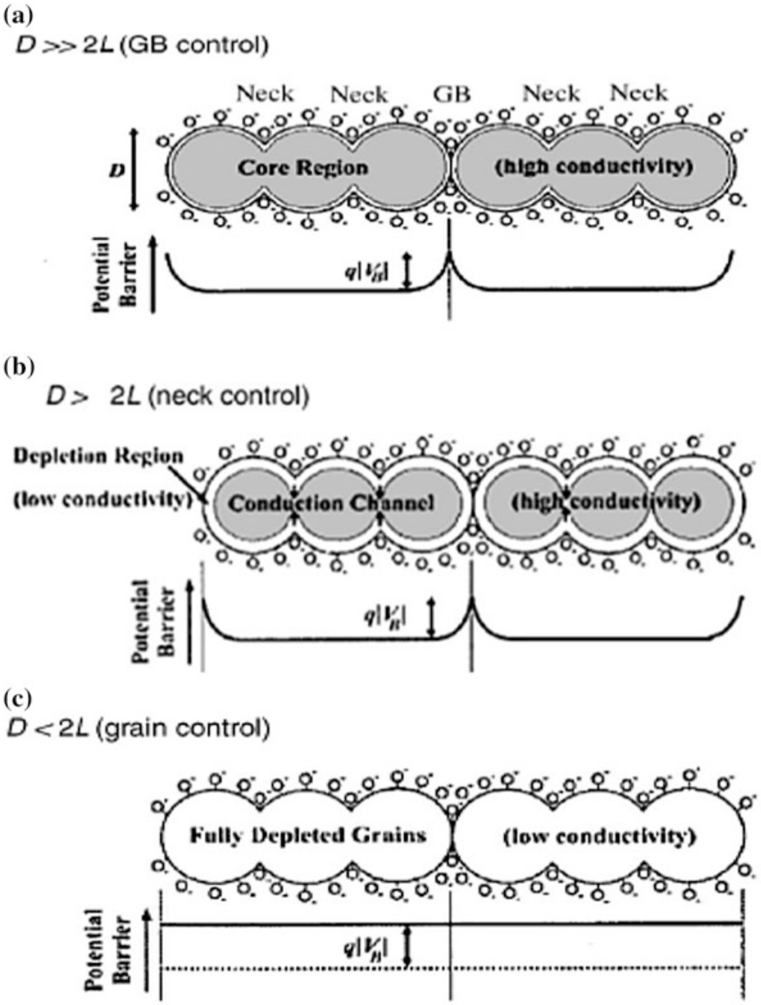


Fig. 8 **a** Grain boundary control, **b** GB + Neck Control, **c** Grain control (reprinted with permission from Ref. [46]) © AIP 2004

the partially sintered crystallites in a metal-oxide film. Those connected crystallites known as grains are further connected to their neighboring grains by regions known as grain boundaries (GB). Xu et al. [31] propose three different cases [6, 7] namely Grain Boundary (GB) Control, Grain Boundary and Neck (GB + Neck) Control and finally Grain Control.

The grain size is taken as D and the depletion layer width or the Debye length [6, 7, 31, 32] is given as

$$L_D = \sqrt{\varepsilon kT/e^2 n}$$

where ε = dielectric constant of the oxide material, k = Boltzmann Const, T = Temperature, e = electronic charge, n = electron density.

Grain Boundary (GB) Control: This occurs when the particle size D is such that the Debye length is negligible compared to it.

In this case the GB barriers (which are modulated by the ambient gas) have the maximum control on charge transfer from one grain to another [6, 32].

GB and Neck Control: For lesser grain sizes, the depletion region occupies into a larger percentage of the volume of grains and the core region (conductive) somewhat shrinks. When D approaches L but is still larger than $2L$, (i.e., when $D > 2L$), the depletion region constricts the conducting paths between adjacent cores.

For such a situation the conductivity depends both on GB and the constricted conducting paths (i.e., necks) between the cores [6, 31, 32].

Grain Control: For grains even smaller in size ($D < 2L$) the grain becomes fully depleted of mobile carriers. In such a condition the inter-crystallite barriers have no significant role to play, but the intra-crystallite barriers control the conductivity of the film.

Now from the above discussion it is understood that a lesser particle size would produce better sensitivity. Studies by Komem and Rothschild on Tin Oxide nanoparticles in the size range 5–80 nm yielded the relationship that the sensitivity S of nanocrystalline metal-oxide gas sensors are dependent on the grain size D as

$$S \propto 1/D$$

Thus it may be concluded from the above-mentioned studies that smaller particle size (and higher surface to volume ratio) will augment the sensitivity in the following manner (i) Increasing the active area of gas sensing in the film (ii) Promoting the surface adsorption of gaseous species by increasing the number of dangling bonds.

This model proposed in 2004 by Rothschild and Komem [32] which is based mostly on computer simulations is indeed supported by earlier experimental data by various researchers. Among them a significant work by Yamazoe et al. [6] clearly illustrates (Fig. 9) the enhancement of sensor response with decreasing particle size of Tin Oxide.

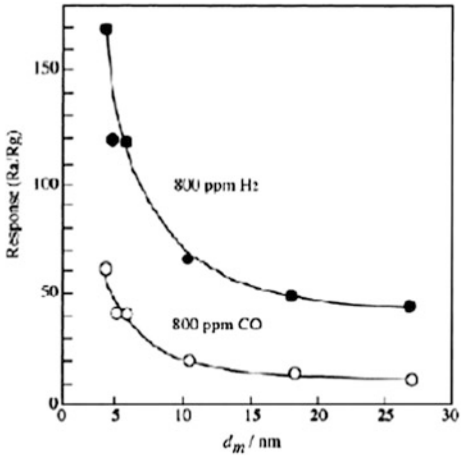


Fig. 9 Effect of grain size reduction on response (reprinted with permission from Ref. [6]) © Springer LLC 2003

13 The Nano Advantage

From the previous section we may concur that application of nanotechnology can play a significant role in the gas sensing properties of thin films. Nanocrystalline thin films with particle sizes of few nanometers show their clear superiority to conventional thick films or pressed pellets in the following sectors.

13.1 Very High Surface to Volume Ratio

It is well known that nanocrystalline materials offer high surface to volume ratio than any other material. With particle size of about 30 nm only 5 % of the atoms are on the surface, as we reduce particle size to 10 nm, we get 30 % atoms and as we go to even smaller dimensions (~3 nm) we achieve 50 % of atoms on the surface. Needless to say that this increases the surface energy exponentially and we get a huge no of dangling bonds on the surface of the material. In this way we increase the effective sensitive area many folds and hence greatly enhance sensitivity. It has also been seen that there exist certain optimum values of particle size for certain target gas-sensing material combination. (i.e., what may be optimum particle size for sensing CO with tin oxide may not be a good choice for hydrogen sensing with Titanium Dioxide).

13.2 Lowering of Operating Temperature

Since nanomaterials have higher numbers of adsorption sites the operating temperature of the sensor is also expected to be less. It has been experimentally seen that for ordinary thick film microcrystalline sensors the operating temperature may be as high as 350 °C but for nanocrystalline materials it can be lowered down to ~200 °C [15–18]. Also notable is the fact that the material shows a much greater response at the same temperature compared to a microcrystalline thin film.

13.3 Precise Control over Sensitivity by Size Reduction

By adjusting particle size, porosity, thickness of the film, in short "Tailoring the material" according to our needs we can increase the sensitivity, selectivity, and response time of our sensor.

14 The Effect of Sensitization

As researchers looked to optimize the performance of sensors, they tried out several new techniques such as the use of catalysts. Historically the catalytic action has been of great significance to surface science and sensors are no exception to that.

In gas sensors Most of the catalysts used are noble metals (Pt and Pd being the most popular among them). This is because of the non reactivity of such metals to most chemicals and gases; and also owing to the fact that they posses good wear resistance [5–7, 32].

Another interesting aspect is the formation of Schottky barriers between certain noble metal–metal oxide interface, which could increase the sensitivity by the mechanism of *Barrier modulation* [8–14]. Nanocrystalline oxide semiconductor surface contains a large number of defects. These defects help in gas-sorption but create barriers against current conduction by acting as the recombination center. Incorporation of noble metals passivate these defect states to a large extent, thus removing the barrier against current conduction [6, 7, 31].

The sensitivity of the gas sensor may be significantly improved by the catalytic properties of the noble metals such as Pd. The dispersed catalytic metal clusters activate the spillover process. These noble metals promote the sensing by lowering the activation energy of the surface reaction. Now the sensitization can be of two types, *Chemical sensitization* and *Electronic sensitization.* Consider a noble metal cluster upon a nanocrystalline oxide thin film. The catalytic/sensitizing action can take place in the following two mechanisms [6, 31–33]:

Chemical Sensitization: The target gas is activated and H is adsorbed on the surface of the noble metal. Now this effect of adsorption is followed by a spill-over of some adsorbed H from the noble metal to the oxide's nanocrystalline surface. The spilled over H then reacts with O^- to give water which immediately desorbs at elevated temperatures (Fig. 10).

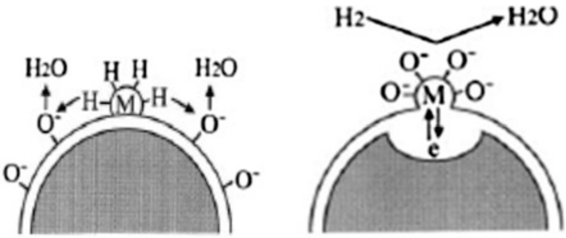


Fig. 10 Chemical and electronic sensitization schematic (*left to right*) (reprinted with permission from Ref. [33]) © Elsevier 1991

The spill-over effect is more pronounced near the edges of the catalyst nano-cluster. Thus, the noble metal promotes H-adsorption; speeding up the reaction mechanism.

Electronic Sensitization: The second type (electronic sensitization), on the other hand, there is no such mass transfer between the additive and the oxide. The additive when oxidized, pulls electrons from the oxide and forms a surface space charge layer, thus reducing conductivity. As the additive is reduced by the target gas it releases the electrons previously captured and thus makes it more conductive. The difference of work function of the metal oxide between the oxidized and reduced state often being large; brings about a considerable increase when introduced to the target gas molecules [6, 31–33].

Even though there exist many theories and proposed mechanisms there are still no concrete experimental proof about exactly how the catalysis mechanism proceeds and there still exist a considerable amount of research yet to be done regarding this.

15 Explosive Gas Sensing with Thin Film Sensors: Methane Sensing as a Case Study

Methane is the simplest alkane, is found in abundance in natural gas (87 % by volume) and is usually present in wetlands as Marsh-gas and especially in Coal mines. It is a colorless, odorless, non-toxic gas but at the same time it is a highly explosive gas (LEL = 5 %). The fact that methane is a colorless, odorless gas, makes it quite difficult to detect for humans and thus even more dangerous. The following sectors require methane monitoring:

(a) LNG/CNG (Liquefied Natural Gas/Compressed Natural Gas) fuel in industry/automobile (b) Industrial/Residential Complexes (especially those built on landfills or low-land areas) (c) Chemical Industry (d) Coal Mines.

The coal mines are probably the most potentially hazardous case of methane related accidents. The danger here comes from the fact that unlike Industry, Automobile, or homes methane here is in an uncontrolled environment and evacuation of personnel from the mining environment is slow and complicated which makes the proper monitoring in such an atmosphere a real challenge. Currently, a high percentage (>50 %) of mining accidents in are related to methane caused explosions. The US standard for Lower Alarm Limit (i.e., the concentration where the alarm should be set off) is (0.3–1.25 %).

Another aspect of methane sensing is the environmental monitoring for The Greenhouse Effect. Methane is a greenhouse gas which can contribute significantly to global warming due to its long (8.5 years) lifetime in the atmosphere [34].

15.1 Methane Detection with ZnO

Many materials (oxides of tin, tungsten, titanium, etc.) have been reported for the application in the field of sensing explosive gases (like hydrogen, methane, propane, butane, etc.) Zinc oxide is a relatively new material for such applications. Given below is a list of such materials for different target gases compiled from a highly cited review paper by Comini [7] which gives an overview of the nanocrystalline materials recently used for gas sensing based on published studies (up to 2006). While mostly SnO_2-based sensors have been used for detecting most of the target gases to date since 2005 there has been some reports on ZnO-based sensors. The use of ZnO for hydrogen sensing using DC resistive technique was first reported in 2005 by Wang et al. [8] (Table 3).

Since hydrogen methane is a reducing gas as well the same principle of sensing H_2 can be applicable for sensing methane with ZnO. For a relatively new sensing material, ZnO has the following advantages:

Table 3 Table of reported oxide materials for gas sensing (Ref. [7]) © Elsevier 2006

Year	Metal oxide	Target species
2002	SnO_2	CO, NO_3, O_2
2003	TiO_2	H_2, CO, NH_3, O_2
	In_2O_3	NH_3, NO_2
2004	Cd: ZnO	RH
	ZnO	Ethanol, O_2
	SnO_2	CO, N_2, O_2
	In_2O_3	Ethanol, NO_2
2005	Ru: SnO_2	CO, NO_2, ethanol
	SnO_2 + CuO	H_2S
	SnO_2 + Pd	O_2
	SnO_2	NH_3, CO, NO_2, DMMP
	ZnO	RH, CO, HCHO, H_2S, NH_3, H_2
	MoO_3	CO, ethanol
	WO_3	NO_2
	In_2O_3	Lipo protein

1. It has good compatibility with Si. This is very important if we want to go for MEMS technology.
2. ZnO thin films can be grown on SiO_2 with considerable ease by various techniques, such as Sol–Gel, CVD, Sputtering, and so on.
3. ZnO thin films are rendered intrinsically an n-type material due to the numerous oxygen vacancies and point defects created during annealing. Since n-type materials are mostly used for sensing reducing gases; so there is no need to dope ZnO to make it n-type.
4. Finally, there exist many standard ways with regard to ZnO to make many novel nanostructures such as nanorods, nano-wires/mesh, nanoflowers, belts, and nano-ribbons to name a few. This gives a very good scope for inquisitive future research work on gas sensing properties of these materials.

16 Zinc Oxide and Its Properties in Brief

Zinc Oxide is a direct band gap ($E_g \sim 3.37$ eV) semiconductor having Wurtzite (Hexagonal) structure at room temperature. ZnO also shows strong piezoelectric and pyroelectric properties useful for mechanical sensors. ZnO is an optically transparent material making it useful for making electrodes. The wide band gap of ZnO makes it useful for short wavelength optoelectronic applications [35–39].

ZnO has intrinsic electron mobility of and hole mobility of 200 $cm^2/(V\ s)$ and 180 $cm^2/(V\ s)$, respectively [37]. Since most of the deposited ZnO thin films have a large number of point defects (such as oxygen vacancy, zinc interstitial, and zinc vacancy, oxygen interstitial, and anti-site oxygen to name a few); they are usually rendered n-type intrinsically [40]. ZnO can be made p-type by doping with group-I elements Li, Na, K, or group-V elements N, P, and As; but the conductivity at room temperature with such dopants is not very good.

16.1 Methane Sensing Mechanism of ZnO and Pd Sensitization

The mechanism related to sensing of methane on non-stoichiometric metal-oxide surface like ZnO involves the stepwise dissociation of methane into methyl groups and hydrogen(adsorbed) [14, 21], which then react with surface adsorbed O_2^- species on the ZnO surface to yield H_2O and CO_2 as the end products releasing electrons back into the ZnO grains. This causes the resistance of the sensor to drop upon exposure to methane. However, when Pd is present it improves the performance in a considerable way. Pd clusters on the ZnO surface help in adsorption of H (formed by dissociation of methane on oxide surface) [14, 41–45]. Pd also acts as a catalyst to lower the temperature for methane dissociation on oxide surface [42–45].

At the same time Pd forms a loosely bound Pd: O on the surface in presence of ambient oxygen which is reduced by hydrogen to produce H_2O and electrons are released in the process [5–7].

Reaction flowchart for methane sensing:

Methane dissociates step-wise on the heated oxide surface

$$\begin{aligned} CH_4 &\rightarrow CH_3(ads) + H(ads) \rightarrow CH_2(ads) + H(ads) \rightarrow CH(ads) + H(ads) \\ &\rightarrow C(ads) + H(ads) \end{aligned} \tag{1}$$

$$Pd_2 + O \rightarrow Pd : O \tag{2}$$

To describe stepwise:

First, the methane gas molecules comes into contact with the ZnO surface at operating temperature ~200 °C. Then it is activated and chemisorbed on the surface as

$$CH_4 \rightarrow (CH_3)_{ads} + H_{ads}$$

Then it is subsequently broken down through subsequent steps ultimately to

$$(CH)_{ads} \rightarrow C_{ads} + H_{ads}$$

These absorbed Carbon and Hydrogen react with the absorbed O^2 already there on the surface to give the end products CO_2 and H_2O which desorbs immediately from the oxide surface. Pd: O on the metal-oxide surface is reduced by hydrogen to produce H_2O; electrons are released.

It is however worth mentioning that depending on surface conditions, temperature, etc. all the methane atoms may or may not break down completely to yield four adsorbed H and one C atom.

16.2 Setup for Sensor Study

For testing different oxide thin films for their gas sensing properties, a typical response measurement/gas sensor characterization setup consists of a testing chamber made of glass. The thin film sample (which is to be used as the sensor) is put into the chamber's uniform heating zone (temperature variation ±1 °C). Fine gold/copper wires are used as the contact leads. IOLAR grade N_2 was used as the carrier gas for ultra-high purity methane (or any other target gas) which are mixed in proportionate ratio in a mixing chamber followed by a mixing coil. Mass flow controllers (MFC) and mass flow meter (MFM) are used to accurately control the mixture concentrations and the flow rates.

A network of coiled tubes known as mixers and small cylinders known as mixing chambers help in making the mixture of different gases homogeneous.

This mixture is finally introduced into the mixing chamber for testing. The gas pressure in the chamber is to be near 1 atm to simulate ambient conditions. For sensor current/resistance measurements the electrodes are connected to a multimeter.

Before starting each testing cycle the entire setup (gas tubing, mixer, coil, and the testing chamber) need to be initially flushed with IOLAR grade N_2 at a sufficient flow rate of >1000SCCM for 10–15 min, thus reducing the presence of any residual air/other gases in the lines. Then after proper mixing of the gases (carrier gas + target) are introduced in the chamber.

The same procedure of flushing, mixing, and feeding is applied for testing different concentrations of the target gas. Resistance/capacitance readings for resistive/capacitive mode operation of the sensor is recorded with the digital multimeter (Fig. 11).

We defined Response magnitude (S) to be:

$$S = \frac{R_a - R_g}{R_a}$$

where R_a indicates the sensor sample resistance in air (i.e., at 0 % target gas concentration) and R_g is the resistance at a given concentration of the target gas (methane). Two other important performance parameters known as Response time (RST) and Recovery time (RCT) are also measured with PC interface of the multimeter. With such simple setup it is possible to study the gas sensing actions of various nanocrystalline thin films of metal oxides.

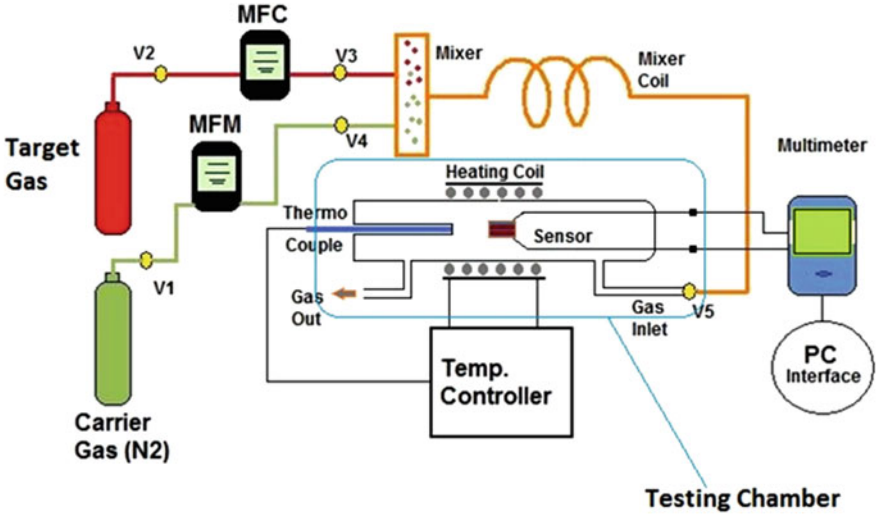


Fig. 11 Schematic of a typical gas sensor setup

References

1. S.M. Sze, *Semiconductor Sensors* (Wiley, New York, 1994)
2. R. Jaaniso, O.K. Tan, *Semiconductor Gas Sensors* (Elsevier, Amsterdam, 2013)
3. *Penguin Dictionary of Electronics*, 4th edn. (Penguin, UK, 2005)
4. A. Sen, Semiconducting oxides in gas sensing. Sci. Cult. **71**, 178–184 (2005)
5. P.T. Mosley, Solid state gas sensors. Meas. Sci. Technol. **8**, 223–237 (1997)
6. N. Yamazoe, G. Sakai, K. Shimanoe, Oxide semiconductor gas sensors. Catal. Surv. Asia **7**, 63 (2003)
7. E. Comini, Metal-oxide nano crystals for gas sensing. Anal. Chim. Acta **568**, 28 (2006)
8. H.T. Wang, B.S. Kang, F. Ren, L.C. Tien, P.W. Sadik, D.P. Norton, S.J. Pearton, J. Lin, Hydrogen-selective sensing at room temperature with ZnO nanorods. Appl. Phys. Lett. **86**, 243503 (2005)
9. P. Bhattacharyya, P.K. Basu, H. Saha, S. Basu, Fast response methane sensor based on Pd (Ag)/ZnO/Zn MIM structure. Sensor Lett. **4**, 1–6 (2006)
10. P. Bhattacharyya, P.K. Basu, H. Saha, S. Basu, Fast response methane sensor using nanocrystalline zinc oxide thin films derived by sol–gel method. Sens. Actuators B **124**, 62–67 (2007)
11. P.K. Basu, S.K. Jana, H. Saha, S. Basu, Low temperature methane sensing by electrochemically grown and surface modified ZnO thin films. Sens. Actuators B **135**, 81–88 (2008)
12. P. Bhattacharyya, P.K. Basu, N. Mukherjee, A. Mondal, H. Saha, S. Basu, Deposition of nanocrystalline ZnO thin films on p-Si by novel galvanic method and application of the heterojunction as methane sensor. J. Mater. Sci. Mater. Electron. **18**, 823–829 (2007)
13. P.K. Basu, P. Bhattacharyya, N. Saha, H. Saha, S. Basu, Methane sensing properties of platinum catalysed nano porous zinc oxide thin films derived by electrochemical anodization. Sensor Lett. **6**, 1–7 (2008)
14. P. Bhattacharyya, P.K. Basu, B. Mondal, H. Saha, A low power MEMS gas sensor based on nanocrystalline ZnO thin films for sensing methane. Microelectron. Reliab. **48**, 1772–1779 (2008)
15. B.S. Kang, Y.W. Heo, L.C. Tien, D.P. Norton, F. Ren, B.P. Gila, S.J. Pearton, Hydrogen and ozone gas sensing using multiple ZnO nanorods. Appl. Phys. A **80**, 1029–1032 (2005)
16. O. Lupan, G. Chai, L. Chow, Fabrication of ZnO nanorod-based hydrogen gas nanosensor. Microelectron. J. **38**, 1211–1216 (2007)
17. H. Gong, J.Q. Hu, J.H. Wang, C.H. Ong, F.R. Zhu, Nano-crystalline Cu-doped ZnO thin film gas sensor for CO. Sens. Actuators B. **115**, 247–251 (2006)
18. M.B. Rahmani, S.H. Keshmiri, M. Shafiei, K. Latham, W. Wlodarski, J. du Plessis, K. Kalantar-Zadeh, Transition from n- to p-type of spray pyrolysis deposited Cu doped ZnO thin films for NO_2 sensing. Sensor Lett. **7**(4), 621–628 (2009)
19. M. Choudhury, S.S. Nath, D. Chakdar, G. Gope, R.K. Nath, Acetone sensing of ZnO quantum dots embedded in polyvinyl alcohol matrix. Adv. Sci. Lett. **3**(1), 6–9 (2010)
20. T.J. Hsueh, S.J. Chang, Highly sensitive ZnO nanowire ethanol sensor with Pd adsorption. Appl. Phys. Lett. **91**, 053111 (2007)
21. V.R. Shinde, T.P. Gujar, C.D. Lokhande, Enhanced response of porous ZnO nanobeads towards LPG: effect of Pd sensitization. Sens. Actuators B. **123**, 701–706 (2007)
22. C.D. Lokhande, A.M. More, J.L. Gunjakar, Microstructure dependent performance of chemically deposited nanocrystalline metal oxide thin films. J. Alloys Comp. **486**, 570–580 (2009)
23. G.C. Yi, C. Wang, W. Park, ZnO nanorods: synthesis, characterization and applications. Semicond. Sci. Technol. **20**, 22–34 (2005)
24. M. Wang, C.H. Ye, Y. Zhang, G. Min, Synthesis of well-aligned ZnO nanorod arrays with high optical property via a low-temperature solution method. J. Crystal Growth **291**, 334–339 (2006)

25. A. Dev, S. Kar, S. Chakrabarti, S. Chaudhuri, Optical and field emission properties of ZnO nanorod arrays synthesized on zinc foils by the solvothermal route. Nanotechnology **17**, 1533–1540 (2006)
26. T.X. Wang, T.J. Lou, Solvothermal synthesis and photoluminescence properties of ZnO nanorods and nanorod assemblies from ZnO nanoparticles. Mater. Lett. **62**, 2329–2331 (2008)
27. S. Dalui, S.N. Das, R.K. Roy, R.N. Gayen, A.K. Pal, Aligned zinc oxide nanorods by hybrid wet chemical route and their field emission properties. Thin Solid Films **516**, 8219–8226 (2008)
28. A.M. Peiro, J.A. Ayllo, J. Peral, X. Domènech, C. Domingo, Microwave activated chemical bath deposition (MW-CBD) of zinc oxide: influence of bath composition and substrate characteristics. J. Cryst. Growth **285**, 6–16 (2005)
29. D. Patranabis, *Sensors and Transducers*, 2nd edn. (PHI, India, 2004)
30. N. Setter, *Electroceramic-Based MEMS: Fabrication-Technology and Application* (Springer, Berlin, 2006)
31. C. Xu, J. Tamaki, N. Miura, N. Yamazoe, Roles of shape and size of component crystals in semiconductor gas sensors. J. Electrochem. Soc. Jpn **58**, 1143 (1990)
32. A. Rothschild, Y. Komem, The effect of grain size on the sensitivity of nanocrystalline metal-oxide gas sensors. J. Appl. Phys. **95** (2004**)**
33. N. Yamazoe, New approaches for improving semiconductor gas sensors. Sens. Actuators B. **5**, 7 (1991)
34. http://en.wikipedia.org/methane.html
35. W.D. Callister, *Materials Science and Engineering* (Wiley , New York, 2007). ISBN:10:81-265-1076-5
36. W.D. Kingery, H.K. Bowen, D.R. Uhlmann, *Introduction to Ceramics*. ISBN: 0471478601
37. S.M. Sze, *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York, 2007). ISBN: 9971-51-266-1
38. Z.L. Wang, *Nanostructures of Zinc Oxide.* Mater. Today, 26–33 (2004)
39. Z.L. Wang, Zinc oxide nanostructures: growth, properties and applications. J. Phys. Condens. Matter **16**, R829–R858 (2004)
40. H.S. Kim, E.S. Jung, W.J. Lee, J.H. Kim, S.O. Ryu, S.Y. Choi, Effects of oxygen concentration on the electrical properties of ZnO films. Ceram. Int. **34**(4), 1097–1101 (2008)
41. H. Yu, Z. Zhang, M. Han, X. Hao, F. Zhu, A general low-temperature route for large-scale fabrication of highly oriented ZnO nanorod/nanotube arrays. J. Am. Chem. Soc. **127**, 2379 (2005)
42. P. Mitra, H.S. Maiti, A wet-chemical process to form palladium oxide sensitiser layer on thin film zinc oxide based LPG sensor. Sens. Actuators B. **97**, 49–58 (2004)
43. R.R. Salunkhe, D.S. Dhawale, U.M. Patil, C.D. Lokhande, Improved response of CdO nanorods towards liquefied petroleum gas (LPG): effect of Pd sensitization. Sens. Actuators B. **136**, 39–44 (2009)
44. A. Giannoudakos, T. Agelakopoulou, I. Asteriadis, M. Kompitsas, F. Roubani-Kalantzopoulou, Development and characterization of ZnO, Au/ZnO and Pd/ZnO thin films through their adsorptive and catalytic properties. J. Chromatogr. A. **1187**, 216–225 (2008)
45. N. Mukherjee, S.F. Ahmed, K.K. Chattopadhyay, A. Mondal, Role of solute and solvent on the deposition of ZnO thin films. Electrochim. Acta **54**, 4015–4024 (2009)
46. A. Rottschid, Y. Komem, J. Appl. Phys. **95**(11) (2004)